Oldenburg Discussion Papers in Economics

Peer Effects in Secondary Education: Evidence from the 2015 Trends in Mathematics and Science Study Based on Homophily

Bernhard C. Dannemann

V – 428-20

February 2020

Department of Economics
University of Oldenburg, D-26111 Oldenburg

Peer Effects in Secondary Education: Evidence from the 2015 Trends in Mathematics and Science Study Based on Homophily *

Bernhard C. Dannemann†

This Version: January 2020

Abstract

In the research on peer effects, unweighted mean classroom performance is the predominant measure used in the estimation of education production functions. In this paper, based on the sociological concept of homophily, I introduce social network matrices that correspond to a weighting scheme for peers in the same class at school. Using spatial regression techniques, I confirm the presence of peer effects for the eighth grade population in the USA in the TIMSS 2015 student assessment. For students, the likelihood of cooperation increases conditionally on visible and non-visible characteristics, such as age, gender, migratory background, and attitudes towards scholastic achievement. This grouping behavior is found to affect the spillover effects of student variables, such as gender and language skills. The main findings are robust to various definitions of the social network matrix, as well as to the inclusion of teacher fixed effects.

Keywords: Human Capital, Cognitive Skills, Peer Effects, Spatial Model, Class Heterogeneity, Education Production Function

JEL Classification Numbers: I21, D62, C31, C18, D91

*I would like to thank my supervisors, Jürgen Bitzer and Erkan Gören, the 2019 seminar participants at the Carl von Ossietzky University Oldenburg, as well as attendees at the annual meeting of the German Economic Association (VfS) 2019, University of Leipzig, for helpful comments and suggestions. All remaining errors are my own.

†Carl von Ossietzky University Oldenburg, School of Computing Science, Business Administration, Economics, and Law (Faculty II), Institute of Economics, Building A5, 26111 Oldenburg, Germany, Tel.: +49-441-798-2625,
website: https://uol.de/en/economics/applied-macroeconomics/team/bernhard-dannemann,
e-mail: bernhard.dannemann@uni-oldenburg.de

1 Introduction

Peers play a pivotal role in educational outcomes according to research showing that the socio-economic composition of school environments is a relevant factor in the acquisition and formation of cognitive skills within and outside of school contexts. Yet while the common determinants of individual student achievement are well researched at the individual level, cooperation and interaction among students in the classroom has frequently been disregarded (Hanushek and Woessmann, 2011). When these so-called peer effects are taken into account at all, they are often implemented as averages of peer performance at the school or grade level, without an appropriate differentiation by class or real peer groups (Angrist, 2014; Lin, 2010). Furthermore, most of the discussion on peer effects has focused on technical problems of identification and estimation while efforts to clarify the modes of effect and transmission channels of peer interaction within classrooms have been left to future research. Data restrictions, such as missing information on friendships and class membership, have led to the use of increasingly large groups of potential peers. With regard to the extant literature on peer effects, this casts doubt on the idea that the relation between peer and individual outcomes is truly existent, and suggests that it is instead merely a correlation that lacks causality. The selection of relevant peers would provide a robust basis to examine the relationship between peer performance and attributes and individual educational outcomes.

Sociological theory suggests a more dynamic way of describing peer interactions by focusing on interpersonal relations. Considering the likelihood of peer interaction and interpersonal ties in closed groups, such as workplaces, school classes or neighborhoods, allows for a description of the social network and the positioning of all its nodes. I hypothesize that within a school class, dyadic peer relations can be postulated based on *status* and *value homophily*, that is, that students group together and cooperate according to both visible and latent characteristics in order to form homogeneous groups (McPherson et al., 2001). Considering this way of forming patterns of interaction within a class makes it possible to determine and assess the importance of other students in the construction of peer averages and to attain a more relevant measure of peer variables even in the absence of actual reported relations (Lin, 2010).

Based on the 2015 Trends In Mathematics and Science Study (TIMSS), I estimate an augmented education production function (EPF) to quantify the magnitude of peer effects in secondary education. I use spatial regression techniques, which aim at modeling geographic interdependencies, together with maximum likelihood estimation to model students in a classroom as a social network in order to identify peer effects. To provide dyadic measures of similarity for students in each class, I follow the concepts of *status* and *value homophily* to calculate different social network matrices. These are employed in the regression to quantify the magnitude of externalities arising in classrooms for $8,387$ students in eighth grade in the United States. This explicit modeling of peer relations facilitates the understanding of how classmates affect each other by introducing a theory-based weighting scheme of relevant peers.

The rich data allows me to control for various individual (e.g., gender, age, migratory background, and frequency of English language use) and parent level characteristics (e.g., education, home resources, and migratory background). By additionally including teacher fixed effects in the regression[1], I can implicitly control for correlated effects and thus address the issue of unobserved heterogeneity on the classroom or subject level. Furthermore, I can separate endogenous and exogenous peer effects, meaning that individual outcomes can be affected by both peer outcomes and peer characteristics (Manski, 1993). I can control for further auto-correlation arising from unobserved common shocks by implementing spatially clustered standard errors at the classroom level. This thorough selection of possible determinants combined with concentration on relevant peers in a classroom as a closed group increases confidence in the peer effect estimates.

In a preliminary step, I use an ordinary least squares (OLS) model or simple linear model (SLM) to estimate the education production function using only standard inputs, such as individual and parental characteristics, as well as teacher fixed effects, to account for school and environmental factors. As a benchmark, I proceed by using a social network matrix that corresponds to the standard linear-in-means (LiM) model for each classroom[2]. These unweighted peer effects show a significant and positive association with individual-level achievement. This applies for both models with endogenous peer effects, as well as models with endogenous and exogenous peer effects.

As a second step, for a more detailed analysis of how and in what group structures students are expected to collaborate, I construct three theory-based similarity matrices to account for different scenarios mimicking the tendency of students to work and cooperate in homogeneous peer groups. The first matrix is based on visible characteristics discussed in the literature (e.g., gender, age, migratory background) and corresponds to the theory of *status homophily*. The second network matrix is based on variables for attitude towards learning and scholastic performance, as suggested by the concept of *value homophily*. As this also encompasses the basic idea of heterogeneous effects in peer relations, this influence depends partly on the students' effort and commitment to achievement. Third, a matrix combining the ideas of *status* and *value homophily* incorporates both theories simultaneously.

In the model specifications that only include endogenous peer effects and thus impose parameter restrictions on the exogenous peer variables, the overall model fit increases considerably relative to the SLM case. The estimated effects of variables for which a classroom externality is expected (e.g., foreign language or perceived bullying) increase in magnitude. However, the different social network matrices produce very similar results, with some changes in the magnitude of the gender achievement gap in the case of *status homophily*. Eliminating the restriction on exogenous peer effects results in clearer changes in the results and leads to more variation across the specifications based on different social network matrices. Introducing grouping behavior mitigates the negative classroom externality originating from a lack of language skills, which is partly offset by a positive effect of migratory background. Additionally, large changes in the magnitude of the gender achievement gap occur, even causing it to lose significance in the case of *status homophily*. These specifications highlight the adverse effect of high age for a given grade on both individual

[1] In the US TIMSS sample, teachers are specific to schools and classes.

[2] More specifically, this corresponds to the leave-me-out mean, that is, the mean performance of all students except for the specific one under study.

and peer achievement. Overall, relying on theory-based social network matrices in the estimation of peer effects in the EPF increases confidence in the appropriate measurement of peer influence by highlighting the role of relevant peers within a school class.

This paper contributes to the literature on the modeling of peer effects, which are of crucial importance to policy discussions on school choice and voucher programs[3], and on tracking programs. Peer effects are also an important topic in research on the permeability of educational domains, particularly in discussions on how to integrate children of refugees into classes of primarily native students (UNESCO, 2019).

The seminal work of Manski (1993) provides a formal description of peer effects and their separation into an endogenous and exogenous effect, as well as the role of correlated factors. The subsequent literature on peer effects has aimed mainly at the correct identification of peer effect parameters. Simpler models employing average class performance as a measure of peer effects are at risk of not being able to separate pure peer effects from correlated effects, that is, common exogenous factors such as socio-economic or institutional environment. Accordingly, a large literature has emerged dealing with approaches to overcome the *reflection problem* described by Manski. However, most of these approaches rest on the properties of the datasets used, which limits the application to a handful of countries and, more importantly, forces one to make crucial assumptions on the relevant peer relations and the level at which they arise, that is, by school, grade, neighborhood, or even higher levels of aggregation. Such broad definitions impede the identification of true causal effects as they incorporate the influence of possibly irrelevant peers (Angrist, 2014).

A handful of studies have chosen to control for peer effects implicitly through the inclusion and interaction of fixed effects in panel data sets (Arcidiacono et al., 2012; Badland and Schofield, 2006). In a setting of stable classes or groups, the unobservable baseline ability of peers can be captured through the inclusion of student-level fixed effects. However, the majority of large-scale student achievement tests, such as TIMSS, which is used here, as well as the Progress in International Reading Literacy Study (PIRLS) and the OECD's Programme for International Student Assessment (PISA) as the most prominent examples, are set up as a repeated cross-section, thus rendering fixed effect estimation techniques inapplicable. Instead, these datasets require an explicit modeling of peer relations, for which the average of peer variables is usually chosen.

Only rather recently has the technique of spatial regression been employed i research on peer effects in various domains, including achievement outcomes, recreational activities, and healthy lifestyle (Lin, 2010; Bramoullé et al., 2009; Boucher et al., 2014; Lee, 2006). Spatial models allow for the consideration of social networks, by permitting the degree to which each peer affects the individual to vary. The structure of the peer group can thus be represented as a matrix of social interactions. Again, the aforementioned studies rely on datasets that offer information on friendship structures, based, for instance, on information reported by students on their closest friends.

This paper contributes to the literature by using sociological theory to construct social network matrices that describe classroom structures in the 2015 TIMSS. These social network matrices can be employed in

[3]See, e.g., the No Child Left Behind (NCLB) act or the Boston Metco program (Angrist and Lang, 2004)

spatial regression approaches to model peer effects in the 2015 TIMSS dataset, even in the absence of explicit information on friendships. This strategy makes it possible to assess smaller groups of relevant peers and thus to resolve doubt as to the causality of positive peer effects.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on adolescents' social interaction and integration into personal networks and highlights the relevance of these factors in the estimation of EPFs. Section 3 outlines the empirical study together with the construction of sociology-based social network matrices and a detailed discussion of the main control variables and their respective transmission channels. Section 4 provides the main results as evidence for the role of peer effects in the estimation of individual-level student performance and a comparison to other studies. Section 5 summarizes the main findings and concludes.

2 Theoretical considerations

2.1 Key challenges in the identification of peer effects

The literature on determinants of cognitive abilities suggests that they should be treated as the outcome of a production process, with standard inputs, such as parent, school, peer, and individual factors (Hanushek and Woessmann, 2011). To model the cross-sectional level of achievement, Todd and Wolpin (2003) suggest using a contemporaneous specification, although it imposes some assumptions and restrictions on the conditioning variables used. First, this specification assumes that only contemporaneous inputs are used in the production of current achievement. Second, inputs are assumed to be constant over time, or at least that their contemporaneous level provides a valid approximation of the history of inputs. Third and last, it is required to assume that the contemporaneous inputs are not correlated to any unobserved variables. These restrictions leave the prior equation reduced to a form incorporating only contemporaneous inputs as portrayed in equation (1), in which achievement specific to individual i in country c at time t is determined by factors such as family F and school inputs S, as well as peer effects P, socio-economic and cultural environment E, and other relevant factors ε (e.g., health and innate abilities).

$$O_{ict} = O_t(F_{ict}, S_{ict}, P_{ict}, E_{ict}) + \varepsilon_{ict} \tag{1}$$

For a valid identification of peer influences, the possible effects of peer outcomes and peer characteristics must be separated. The literature on peer effects has adopted the vocabulary of Manski (1993), who distinguished peer effects by the context in which they arise. First, in *endogenous effects*, the behavior of the individual studied is affected by the behavior of related peers. In this case, the outcome under study is affected by the outcome of the peers, that is, individual achievement is partly determined by peer achievement. Second, *contextual* or *exogenous effects* measure how exogenous group characteristics affect the way the individual studied behaves. In the EPF, this would correspond to how the inputs of peers (e.g., individual peer characteristics but not the outcome) affect the outcome of the individual under study. Third and last, *correlated effects* are the cause of individuals behaving similarly as a consequence of having similar characteristics or being exposed the same environments. These are influences that affect all individuals within a school class, for example, school characteristics or neighborhood effects. Depending on the setting, either peer outcomes as an endogenous effect or peer characteristics as an exogenous effect (or potentially a combination of both) could affect individual outcomes. These considerations of transmission channels are crucial for the correct choice of models in the subsequent study. Individual scholastic achievement is partially determined by the performance or behavior of other students in the same school or class or in the same group of friends. In economic terms, these peer effects can best be described as classroom externalities in the production of cognitive skills.

2.2 Definition and measurement of peer groups

Why are peer effects especially relevant in determining the performance of teenagers in school in the first place? Generally, students in secondary education are regarded as vulnerable to both positive and negative influences, their character and personality are still being formed (Steinberg and Monahan, 2007). In this study, students from the US TIMSS population, with an average age of about 14.2 years, are one of the most susceptible age groups to behavioral influence from peers both in and outside of school contexts. Sorensen et al. (2017) analyze specifically how the influence of peers and families evolves over time. Their main finding suggests that role models shift from fourth to seventh grade. The importance of parental influence in educational achievement disappears almost completely, while that of peer influences increases significantly. This shift is attributed mainly to a change in role models with age: For teenagers, it is primarily peers who determine attitudes and effort towards learning and the allocation of time to different activities or hobbies.
Peer effects are generally operationalized by choosing a peer group and taking the appropriate averages of the relevant variables. Simple as it may sound, this raises three questions. First, how is a peer group defined and which individuals should be considered as a part of it? Second, what is the appropriate average and how can the relative importance of peers be assessed? Third, which variables of the potential peers are indeed relevant?

The first question is perhaps the most difficult one to answer, as influences on individual students can come from a multitude of potential peers, including students in the same grade, students in others grades, and even students at other schools. If one includes interactions online (e.g., on online gaming platforms or social media), virtually anyone could qualify as a member of the peer group. Thus, a prerequisite is the definition of clear-cut criteria for peer group membership. Considering peer groups by classes means assigning individuals to groups based on a visible, exogenous group membership.[4] Within a school, a class is the smallest closed unit in which students interact and cooperate. They are exposed to a variety of influences by common factors, for example, the teacher, equipment with educational material, or overall performance. Based on this criterion, each class constitutes the population from which an individual draws

[4] Of course, many other definitions of peer groups are conceivable, e.g., by non-observable group memberships (e.g., friendships Lee, 2006) or by informal group characteristics (e.g., common leisure activities Knifsend et al., 2018).

the members of his or her peer group. This ensures that peer effects measure actual relations that fulfill the criterion of real exposure or *propinquity*, that us, geographical proximity (Moreland and Beach, 1992), thus indicating an actual impact of possible peers (as opposed to the concern of selected peers being irrelevant, as addressed in Angrist, 2014). However, studying peer effects on the classroom level requires a thorough consideration of these correlated factors in order to separate the peer effects from the common environmental factors.

The discussion of peer groups then leads to the second question: How students are in fact grouped within the class and how they choose to cooperate. It is thus conceivable, that some students have higher relative importance to an individual than others. Groups could be defined by teachers ex ante to optimize the learning environment, or students could self-select into groups or form their own groups or cliques – or both (Farmer et al., 2018; Freeman et al., 2017). When able to group together freely, students are observed to form groups based around prior achievement, ability, or interest in the topic. Thus the stereotype of the diligent 'nerd' listening attentively in the front row and the underachieving 'lazybones' slouching in the back row applies to a certain degree. Furthermore, gender and socio-economic or ethnic characteristics are found to contribute to the early decision to form a group (Shrum et al., 1988). Undergraduate college students have furthermore been observed to group and collaborate based on prominently visible characteristics such as ethnicity and gender (i.e., to some extent driven by stereotypical threats or the desire to avoid isolation in class), but also by academic history in terms of grade point average or attendance of the same classes (Freeman et al., 2017). This tendency to join or work in relatively homogeneous groups indicates that the sociological theory of *homophily* applies to classroom settings and extends to other environments out the school context (Louch, 2000; Knifsend et al., 2018; Currarini et al., 2009).

However, to establish a proper measure of peer effects, it is necessary to identify the relevant peers who influence an invidual's behavior. The majority of studies fail to do so (Angrist, 2014), either because they attribute the same influence to all peers (e.g., in the linear-in-means model, Manski, 1993) or make unrealistic *ad hoc* assumptions on peer relations Bramoullé et al. (e.g., as described in 2009). Only a few studies have provided information on friendships reported by individuals under examination (e.g. Lin, 2010; Bramoullé et al., 2009), but these used restrictive datasets and suffered from missing information on peer relations for a large fraction of the sampled population.

In spite of the vast existing literature on peer effects, the various channels and directions of these effects are still unclear. Ties between students could serve as channels for the transmission of material and immaterial goods or information, such as help with assignments, studying together in groups, or simply working together in class. There are numerous proponents of positive externalities in the literature who provide examples of how high-achieving peers can boost an individuals performance, for example, by providing encouragement and a beneficial learning atmosphere and creating incentives to reach a similar level of achievement (Hanushek et al., 2003; Burke and Sass, 2013; Carrell et al., 2009; Arcidiacono et al., 2012).

This suggests that efforts to alter school composition (e.g., through vouchers or bus transfers) could serve to improve overall educational performance. Yet, some peer characteristics could have adverse effects, reducing individual performance. One possible negative effect is overall low classroom performance, which could slow down the learning progress of the entire class and thus reduce the efficacy of teaching. However, this is simply the reverse side of the coin from the aforementioned relationship. High achieving peers could also lead to the teacher having higher expectations, thus grading the bottom end of the achievement distribution more strictly (Kiss, 2017). High-achieving peers could also be intimidating to their low-achieving counterparts and thus inhibit learning efforts (Winston and Zimmerman, 2004). These findings speak in favor of selecting peer outcomes, i.e., the endogenous peer effect, as one relevant peer measure in the subsequent analyis.

Additionally, some behavioral peer characteristics could affect individual achievement more than peers' scholastic performance. Several studies find significant and negative impacts of peers who show disruptive behavioral symptoms, such as absenteeism, distraction, or substance abuse. These undesirable traits reduce the effectiveness of the learning environment and furthermore push peers to show similar patterns of disruptive behavior (Grygiel et al., 2018; Carrell and Hoekstra, 2010; Gaviria and Raphael, 2001). These findings illustrate that peer effects can arise in the form of peer characteristics, thus suggesting that for the precise modeling of peer effects, exogenous peer effects should be considered as part of the model as well.

The magnitude of peer effects is likely to depend on the setting, that is, class or school size, and on other factors, for example, socio-economic or cultural background (Carrell and Hoekstra, 2010). The magnitude of the effect also depends on the position in the distribution of peer group achievement. For instance, high-achieving students are found to receive lower benefits from average peer performance than their low-achieving classmates do (Hanushek et al., 2003). Furthermore, peer effects are likely to influence individual achievement even outside school, that is, through common leisure activities, working together on homework, or spending time together in other ways (Knifsend et al., 2018).

2.3 Assessing similarity and social network matrices

Classroom adjacency matrix Social interactions within a class can be portrayed in a social network matrix, in which the individual elements correspond to the dyadic relations between the students. As a precondition, to ensure that individuals are only affected by students of the same class, a binary contiguity social network matrix, the adjacency matrix, is generated. Following this basic idea, two students, i and j, who are enrolled in the same class are interpreted as peers, so that the corresponding element of the matrix, w_{ij}, is set to 1. Of course, when one class is considered, all class members are peers all others, except themselves, so that $w_{ii} = 0$. Accordingly, the classroom spatial weights matrix is a symmetric square matrix constructed as shown in equation (2).

$$\boldsymbol{W}^{(Adj)} = (w_{ij})_{n\times n}, \text{ where } w_{ij} = \begin{cases} 1 & if\ c_i = c_j \\ 0 & if\ c_i \neq c_j \\ 0 & if\ i = j \end{cases} \tag{2}$$

This social network matrix is then row normalized, that is, each row is divided by the number of its non-zero elements. Mathematically, multiplying a variable with this row-normalized matrix corresponds to taking the equally weighted average[5], i.e., $\boldsymbol{W}^{(Adj)} \cdot X = \frac{1}{N}\sum_{i=1}^{N} X_i = \overline{X}$, across all potential peers, which is consistent with a traditional modeling of within-group peer effects. In terms of the Manski (1993) vocabulary, this corresponds to a linear-in-means model, where the individual outcome is determined by the mean achievement and inputs of the peer group.

Similarity in dyadic peer relations To quantify the strength of dyadic peer relations instead of just analyzing a closed group average with uniform importance of peers, it is necessary to evaluate the proximity of all pairs of individuals within a class. The geographic proximity through common class membership could be a driver of the bonding process of potential peers (Moreland and Beach, 1992). Following the tendency to work in homogeneous groups, students are expected to cooperate based on individual level characteristics (gender, age, origin) and attributes (attitude to scholastic performance).

Similarity is measured based on a coefficient by Gower (1971), which is designed for mixed measurement scales of variables[6]. Consistent with the concept of *homophily*, Gower similarity indicates the degree of similarity in a dyadic relationship. Accordingly, all dyadic relationships portrayed in the matrix are symmetric, that is, $S_{ijc} = S_{jic}$. In equation (3), S_{ijc} is the total similarity of all employed attributes $k = \{1, \ldots, \nu\}$ for students i and j who are members of class c.

$$S_{ijc} = \sum_{k=1}^{\nu} s_{kijc}\delta_{kijc} \Big/ \sum_{k=1}^{\nu} \delta_{kijc} \tag{3}$$

To incorporate mixed measurement scales, the computation of s_{kijc} differs according to the variable scale at hand. For quantitative variables, the similarity measure is computed as given in equation (4), where R_{kc} is the range of the variable considered for class c. The parameter δ_{kijc} represents the value 1 when a character k can be compared and is set to 0 otherwise. The formula implies that for each character, the maximum similarity is 1 for the case that $x_{kic} = x_{kjc}$ and the minimum similarity is always zero between the lowest and the highest value of the specific character in each class.

$$s_{kijc} = 1 - |x_{kic} - x_{kjc}|/R_{kc} \tag{4}$$

In the case of dichotomous variables, s_{kijc} and δ_{kijc} are defined according to the rule presented in (5). A similar approach is used for qualitative data, where $s_{kijc} = 1$ if individuals i and j in class c agree on the k-th character and is set to zero otherwise.

$$s_{ijk} = \begin{cases} 1 \wedge \delta_{kijc} = 1 & \text{if } x_{kic} = x_{kjc} = 1 \\ 0 \wedge \delta_{kijc} = 1 & \text{if } x_{kic} \neq kjc \\ 0 \wedge \delta_{kijc} = 0 & \text{otherwise} \end{cases} \tag{5}$$

The modeling of peer relations without using actual observed interpersonal relationships entails five restricting

[5]Omitting the individual under study from the calculation of peer averages is also labeled as the *leave-out* or *leave-me-out* mean (Feld and Zölitz, 2016).

[6]While *age* is a metric variable, both *gender* and *migration background* are binary variables.

assumptions on the social network characteristics. While the first three apply especially to the adjacency matrix, the homophily-based social network matrices are not able to fully alleviate these restrictions.

First, for network density within a classroom, no maximum number of connections or relationships exists. All students in the class are expected to affect each other, although for the case of the similarity based social network matrices, less similar students have a smaller effect and vice versa. Second, no closed clusters exist within classes, which implies that no formation of closed cliques or groups occurs, that is, the class is not split into sub-groups as the result of homophily-related grouping behavior. Third, the multiplexity of the relationship cannot be assessed. Thus, it is not possible to determine at what level a relationship occurs (e.g., friendship, comradeship, or strategic cooperation). Fourth, and applying only to the homophily-based matrices, transitivity in dyadic relations[7] is disregarded, as similarities are determined ex post on objective criteria. Fifth, and last, the constructed similarity measures are symmetric and strictly positive, that is, they correspond to the description of an undirected network, where the direction of influence between two nodes is disregarded. Furthermore, all peer relationships are beneficial, so that the possibility of enmity between students is excluded by definition.

Compared to the baseline adjacency matrix described above, the similarity measures based on the Gower coefficient introduce a weighting of peer influences. Depending on the specification, individuals are more strongly influenced by peers with similar characteristics. Row-normalizing the social network matrices by dividing by the row-wise sum of elements corresponds mathematically to weighting the peer influences by their relative similarity, making the similarity-based social network matrices an extension of the adjacency matrix.

$$\boldsymbol{W} \cdot \boldsymbol{X} = \begin{pmatrix} \frac{1}{\sum_{j=1}^{n} S_{1j}} \sum_{j=1}^{n} S_{1j} \cdot X_j \\ \vdots \\ \frac{1}{\sum_{j=1}^{n} S_{nj}} \sum_{j=1}^{n} S_{nj} \cdot X_j \end{pmatrix} \in \mathbb{R}^{n \times 1} \tag{6}$$

Writing the similarities as the social network matrix $\mathbf{W} := (S_{ij})_{n \times n}$ and any variable X as the vector $\mathbf{X} = (X_1, \ldots, X_n)'$ of in total n classmates, Equation (6) underlines that multiplying the similarity matrix $\mathbf{W}$ by any variable $\mathbf{X}$ is equivalent to calculating the weighted average of the n classmates, using their similarities as weights, resulting in an $n \times 1$ vector of weighted leave-me-out means of the corresponding variable.

Status homophily-based similarity. The similarity measure above can be calculated based on sets of variables, in order to incorporate different forms of homophily into the weighting scheme. The first case is *status homophily*, that is, visible factors that can be considered as fixed. To express the tendency of students to form peer relations according to visible characteristics, I construct the first similarity matrix based on the student's gender, age and nationality (Carrell and Hoekstra, 2010; Diette and Uwaifo Oyelere, 2017). This setting implies that peer groups emerge soon after students are assigned to classes. Students self-select into

[7] For instance, the strength of the relationships between students a and b ($a \leftrightarrow b$), and b and c ($b \leftrightarrow c$) does not affect the strength of the relationship between students a and c ($a \leftrightarrow c$)

groups within the class based on relatively superficial characteristics, with the possible aim of reducing the individual's risk of becoming an outcast (Freeman et al., 2017; McPherson et al., 2001).

Value homophily-based similarity. While the prior matrix is based on visual characteristics and corresponds to a rather short-term formation of groups, it is possible to exploit non-visual characteristics for the evaluation of similarity. Hanushek et al. (2003) list educational achievement in addition to migrant background and socio-economic status as the three most commonly used characteristics in the separation of peer group effects.

The dyadic similarity for the case of *value homophily* is thus assessed based on three factors: first, the student's sense of school belonging; second, scales for the student's confidence in the test subjects mathematics and science; and third, how the students value the two test subjects to evaluate their importance. Including attitude towards educational achievement as a measure on which peer relations are chosen serves two purposes. First, students have been observed to group according to academic performance (Freeman et al., 2017), and second, prior results suggest that peer effects are heterogeneous for different positions in the achievement or skill distribution (Hanushek et al., 2003; Sorensen et al., 2017). Accordingly, the *value homophily* social network matrix can be interpreted as a rather long-run formation of peer groups, where students have demonstrated or proven their valuation of academic performance.

While *status homophily* is interpreted as a short-run criterion for group formation and *value homophily* corresponds to a long-run result, it is conceivable that a combination of both concepts exists. After an initial grouping based on visual characteristics, students would then start to adjust groups according to the valuation of academic and scholastic performance. Thus, the combined *status* and *value homophily*-based social network takes the arithmetic mean of the individual similarity measures of the prior matrices into account to portray a consideration of both theoretical approaches.

A numerical example. Figure 1 provides and visualizes a numerical example of how dyadic similarities between students are calculated based on Gower's similarity measure. The example is based on a class of four students with three attributes (gender, age, and attitude) as listed in panel (a).[8]

Applying equation (3) to the data yields the dyadic similarities s_{ijk} between individuals i and j on variable k, as presented in panel (b). The total similarity S_{ij} corresponds to the mean of the individual similarities. The similarity of individuals to themselves (i.e., $i = j$) is, by definition, set to zero.

A visualization of the difference between the adjacency and the similarity-based social network matrices is made in panels (c) and (d). Panel (c) corresponds to the adjacency matrix, where all four students are assigned a uniform weight, regardless of their (dis)similarities on the variables *age*, *gender* and *attribute*. In contrast, panel (d) incorporates the dyadic similarities on the aforementioned variables, as is shown by the corresponding numbers and indicated by the line thickness. For example, the similarity between students 1

[8]For the sake of simplicity, the number of variables employed in the calculation of dyadic similarities is reduced and does not correspond to actual observed individuals from the original sample.

and 4 ($S_{1,4} = 0.67$) is the highest within the class and is almost twice as high as the similarities between, for example, 1 and 2 ($S_{1,2} = 0.35$) and 1 and 3 ($S_{1,3} = 0.38$), rendering student 4 the most important peer to student i and vice versa.

3 Empirical approach

3.1 Data

The study uses data on US eighth-grade students' achievement from the current 2015 wave of TIMSS conducted by the International Association for the Evaluation of Educational Achievement (IEA). The TIMSS data provide information on the school and class membership of each individual, making it possible to group students by class, instead of eliciting peer relations from grade-level, school, or city populations. Each class constitutes the population from which individuals draw the members of their peer group. This data structure ensures that peer effects measure actual relations which fulfill the criterion of real exposure or *propinquity*, that is., geographical proximity (Moreland and Beach, 1992), thus indicating a real impact of these possible peers (as opposed to the concerns addressed in Angrist, 2014).

TIMSS encompasses schools of all types (public and private), as well as all common educational tracks in the US education system. The students in the estimation sample have an average age of 14.23 years and consist of 50.52% girls and 49.48% boys. 94.50% of the students have stated that they were born in the territory of the United States.[9] In total, the dataset provides information on $8,387$ individuals. Detailed summary statistics are presented in Tables 2 and 5, whereas pairwise correlations are shown in Table 3 in the Appendix. An important feature of the TIMSS dataset is that information is available on how the students are distributed among the 213 schools and 457 classes in the sample. Compared to prior studies, this allows for credible class sizes and thus reasonable peer groups. The sample comprises 7 isolated individuals, for whom no peers could be assigned as they have a unique school and/or class ID. In total, 62 individuals (0.74%) are reported as being in classes with 5 or less students, representing 0.69 per-cent of the sample population. The minimum, mean, median, and maximum class size are 1, 18.7, 20, and 37, respectively, whereas the distribution is shown in Figure 2. [10] Further, information on a total of 836 teachers in the test subjects of mathematics and science is available. [11]

[9]According to supplement 2 of the TIMSS dataset, "United States" includes the 50 states, its territories, the District of Columbia, and U.S. military bases abroad.

[10]In contrast, the sample used in the study based on reported friendship-relations by Lin (2010), contains $68,131$ observations, of which $18,572$ are isolated. These are individuals who have not been reported as a friend by any other person in the data and have also not reported any friends from the dataset. As no information on class membership is used for peer group formation, these individuals are not exposed to peer effects by definition. For the remaining observations, the minimum, mean and maximum group size are 2, 38.5, and 427 observations, respectively.

[11]The majority of students (66.4%) are taught by two teachers. For the remaining students, 20.56% have three, 2.06% one, 7.94% four, and 3.00% five assigned teachers.

3.2 Estimation

In order to evaluate individual student achievement, I estimate an EPF as specified in equation (7). To consider the effect of peers and classroom structure, peer effects are explicitly modeled in the regression using spatial regression techniques based on different social network matrices to represent the sociological concepts of *status* and *value homophily* in the formation of social groups within a classroom.

$$\boldsymbol{o} = \alpha\boldsymbol{\iota}_n + \rho\boldsymbol{W^{(m)}}\boldsymbol{o} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W^{(m)}}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{P}\boldsymbol{\delta} + \boldsymbol{T}\boldsymbol{\gamma} + \boldsymbol{u} \tag{7}$$
$$\text{where } \boldsymbol{u} = \lambda\boldsymbol{W^{(Adj)}}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

Here, $\boldsymbol{o}$ represents the vector of educational outcomes, specific to student i in school s and class c. The expression $\alpha\boldsymbol{\iota}_n$ denotes the intercept. $\boldsymbol{X}$ refers to a matrix of individual characteristics, with the coefficient vector $\boldsymbol{\beta}$, while the matrix $\boldsymbol{P}$ with coefficient vector $\boldsymbol{\delta}$ measures parent-level inputs in the education production process that are, of course, specific to the individual. To account implicitly for the school and classroom environment, that is, the *correlated* effects, a matrix of teacher specific dummies ($\boldsymbol{T}$) with coefficient vector $\boldsymbol{\gamma}$ is included. Note that within the TIMSS USA dataset, teachers are specific to one class. While each class usually has two teachers, that is one science and one mathematics teacher, cases of more than one teacher per test subject are observed.

Additionally, $\boldsymbol{W^{(m)}}\boldsymbol{o}$ refers to the influence of classmates on individual performance, where $\boldsymbol{W^{(m)}}$ is a social network matrix (with m being one of the aforementioned matrices), which can have various definitions (see above), and $\boldsymbol{o}$ is the vector of student achievement. The coefficient ρ corresponds to the *endogenous* peer effect. Similarly, $\boldsymbol{W^{(m)}}\boldsymbol{X}$ refers to spillover effects of peer characteristics on the individual, where $\boldsymbol{X}$ denotes the matrix of student characteristics. The coefficient vector $\boldsymbol{\theta}$ corresponds to the *exogenous effects* described above. Finally, $\boldsymbol{u}$ is the vector of residuals, which are correlated across students in the same class, and $\boldsymbol{\varepsilon}$ is a vector of stochastic error terms. The motivation for modeling dependencies in the error term among peers within a class is to capture unobserved common shocks or factors. Thus, all specifications include auto-correlation in the error term across peers, that is, $\lambda \neq 0$.

Equation (7) allows for the estimation of various settings in which peer effects could arise by imposing parameter restrictions on the scalar ρ and the coefficient vector $\boldsymbol{\theta}$. First, setting ρ and $\boldsymbol{\theta}$ to zero assumes that no peer effects are at work, which is referred to as the simple linear model (SLM). Here, no spillovers arise and the individual student is not affected by peer outcomes or characteristics. This corresponds to the baseline scenario. Second, relaxing the restriction on ρ, but keeping $\boldsymbol{\theta} = \vec{0}$ corresponds to the modeling of an endogenous peer effect only, which allows for spillovers caused by peer outcomes. Of course, the effect of peer attributes on individual achievement affects the individual as well, although by definition with a constant ratio.[12] Third, allowing ρ and $\boldsymbol{\theta}$ to differ from zero corresponds to a combination of endogenous and exogenous peer effects, which is most flexible, as it allows for spillovers to vary both in sign and ratio.

[12] That is, the spillover caused by a peer attribute is predetermined by the effect of the attribute on the peer and the endogenous peer effect parameter ρ.

As Manski (1993) and Lyle (2007) add for consideration, OLS estimation of peer effects is likely to be biased due to the *reflection* problem induced by simultaneity[13]. Technically, using the contemporaneous average achievement of peers as an explanatory variable corresponds to an auto-regressive component across individuals. OLS estimation of this relationship would introduce a spatial dependency into the error term. This leads to a biased coefficient of peer achievement and henceforth to biased estimates of all other coefficients as well. An alternative estimator that is capable of considering the network interdependencies in the dependent variable of students within a class is maximum likelihood (ML) estimation.

3.3 Variable selection

Main dependent variable. For the measurement of student performance outcomes, student level test scores in TIMSS 2015 are used. TIMSS is designed as a paper-based student assessment, conducted in the United States in English. In order to assess a broad domain of skills, the average of the test scores in the subjects of mathematics and science is employed. Overall, student test scores have been found to be highly correlated with results from IQ tests, thus providing a measure of general cognitive ability (Rindermann, 2007).

3.3.1 Student characteristics

Gender. First, the reported sex of each student is included to control for gender-specific differences in classroom performance and attitudes towards schooling. Reportedly, male and female students differ in terms of diligence, ability, and achievement, depending on the subject considered, presumably due to differences in upbringing and cultural background (Guiso et al., 2008; Fryer and Levitt, 2010; Aesaert and van Braak, 2015) or due to gender-specific curricular needs, for example, physical activity as a balancing factor for boys to increase their ability to concentrate and stay focused (Cöster et al., 2018).

Age. As TIMSS is designed as a grade specific study and the focus is on the eighth grade, the age of the observed students is not fixed.[14] Even though there is a mandatory minimum age for enrollment in many US states, parents have some leeway in when they enroll their children (with required school entry between 5 to 7 years), which causes significant age differences ub first-graders with diverse effects on educational outcomes (Bedard and Dhuey, 2006). Students could be younger than others in their grade due to *acceleration*, that is, early enrollment or skipping an entire grade. A young age for grade could thus be caused by high initial abilities, but could also indicate low maturity, cognitive development, or experience, leading to lower classroom performance or social outcomes in later life. This has been the subject of heated discussion in the literature (McEwan and Shapiro, 2007; McCrary and Royer, 2011; Warne and Liu, 2017; Warne, 2017; Kretschmann et al., 2016). In contrast, high age relative to grade, could be the consequence of either voluntary late

[13] That is, mean class performance affects individual performance and vice versa

[14] It differs in this respect from, e.g., the OECD's PISA study, which is targeted at a 15 year old student population which is heterogeneous in terms of grade

enrollment or of grade repetition. The former is thought to improve performance through gains in maturity (Bedard and Dhuey, 2006; McEwan and Shapiro, 2007). Typical reasons for the latter are bad grades, lack of motivation or participation, or absence due to severe illness or other major issues (Hughes et al., 2017).

Migratory background. Additionally, a student's immigrant background is included to account for possible difficulties in the learning environment. This is measured by the student's migration status (i.e., first or second generation immigrant) and the frequency with which the English language is spoken at home. Several studies have found a migratory background to significantly interfere with the academic career. The effect is not exclusively caused by differences in socio-economic background, but also reinforced by the characteristics of the educational system in the country of residence (Borgna and Contini, 2014). For instance, in a selective public tracking system, immigrant children and children from ethnic minorities have been observed to be hampered in the transition to secondary education (Jackson et al., 2012) and are more frequently assigned to non-prestigious educational tracks (Cebolla-Boada, 2011). However, immigrants are found to make more ambitious decisions in their educational careers than their native counterparts, specifically in tertiary education (Kristen et al., 2008). Furthermore, it has been observed that the performance of natives is negatively affected by the presence of immigrant students in class. Thus, being foreign might not only affect the individual but also create classroom externalities (Brunello and Rocco, 2013; Diette and Uwaifo Oyelere, 2017).

Perceived bullying. To account for the effect of harassment or bullying by classmates on the respective student's achievement, an item for perceived bullying activity is included. Bullying has been identified as a determinant of student achievement that has significant negative effects on the outcome, regardless of the grade level or the country (Ponzo, 2013; Resende Oliveira et al., 2018). However, causality could run in both directions, such that high-performing or ambitious students could even face increased bullying (Bekiari et al., 2017; Harel-Fisch et al., 2011).

3.3.2 Family background

Educational achievement. In order to measure the overall socio-economic status as well as the cognitive abilities of the individual's mother and father, the highest educational attainment is included to provide a measure of initial or innate ability. The socioeconomic status and educational level of parents are generally agreed to have a positive correlation with the school performance of their children (Martins and Veiga, 2010), with child health as a possible transmission channel (Currie, 2009). Also, a higher level of education is assumed to increase the parents' valuation of and involvement in their child's education (Rindermann, 2007; Fan and Chen, 2001; Davis-Keane, 2005). Additionally, a strong intergenerational transmission of intelligence is observed: Parent cognitive ability is found to be a strong predictor of the child's educational outcomes (Anger and Heineck, 2010).

Immigration background. Similar to the migrant status of the child, a variable indicating whether each parent was born in United Sates is included. This makes it possible to identify cases in which child born in the

United States has foreign-born parents (either one or both). Accordingly, an approximation of the family's migration history can be used to determine whether the child is a first- or second-generation immigrant.

Proxies for home economic and educational resources. The TIMSS dataset does not explicitly contain a measure of the parents' available economic resources (e.g., income, wealth, real estate), which could affect the child's educational outcomes (Martins and Veiga, 2010). I therefore proxy for resources by the presence of items specific to education or wealth in the home. For instance, the number of books in the home is described as a measure of parenting, home education, priority of knowledge, and social background (Hanushek and Woessmann, 2011) and is found to be robustly associated with student achievement (Schuetz et al., 2008; Freeman et al., 2010). Using a similar interpretation, I include the number of digital devices in the home. The supplement to the TIMSS dataset lists these as a computer or tablet, Internet connection, own mobile phone, or a gaming system. These items are thought to proxy for wealth, but also for adjustment to modern technology, that is, whether the parents have adapted their lifestyle as well as their parenting style to the requirements of technology-rich environments (Hanushek et al., 2015) and enable their children to do so as well. As a measure of family resources devoted to their children, dummy variables for the child having their own desk, own room, or both, are included. This measures the basic prerequisites for the child to be able to review and study school material at home. Especially for adolescents from low-income families, the role of parenting style has been emphasized in the research (Johnsen et al., 2018).

4 Main empirical results

4.1 Endogenous peer effects

The baseline regression results are shown in Table 4, and the partitioned effects are presented in Table 5. Column (1) shows the result for the simple linear model (SLM), which includes neither endogenous nor exogenous peer effects. This reduced model explains approximately 10.6% of the variation in test scores, in addition to the teacher fixed effects, which were partialled out in a preliminary regression. Consistent with existing literature, a positive and significant gender achievement gap of 12.2406 points is found. An additional year of age for the individual under study is associated with a 11.3951-point decrease in test scores. Interestingly, none of the variables measuring the frequency of language use enters the regression model significantly. For perceived bullying, a positive and significant impact is found, which appears puzzling at first. However, this might be driven by reverse causality, that is, being bullied because of high classroom performance. The columns (2-5) refer to the SAC model, which only includes endogenous peer effects, that is, peer achievement, but imposes the restriction of no exogenous peer effects, that is, no influence of peer attributes. This implicitly assumes that peer characteristics have no influence on individual achievement other than their determining effect on peer achievement.

Column (2) uses the classroom adjacency network matrix to model endogenous peer effects and therefore corresponds to the LiM model. The coefficient of the endogenous peer effect is positive and significant at the 1% level and amounts to 0.7476. The magnitude of externalities within the classroom can then be assessed based on the effect decomposition, which indicates that the indirect effects are approximately 2.5 times as large[15] as the direct effects including feedback mechanisms. In this specification, the direct effect of being a male student is significantly positive at 13.686 test score points. The indirect effect, that is, the hypothetical scenario of all other students being switched from female to male, would correspond to a 34.652-point increase in the individual's test performance, again significant at the 1% level. The effect of age is estimated as negative, with a total effect of -52.070 points, significant at the 1% level, thus supporting the hypothesis of high age for grade being caused mainly by grade repetition and not late enrollment. The frequency of language use at home of the individual student does not have a clear effect on student achievement in this model specification. While the effect of only *sometimes* speaking the English language at home is negative and significant at the 10% level, the other categories do not enter significantly. This also applies to the indicator variable for being born outside of the United States, which is not significant at conventional levels. Last, the effect of perceived bullying is again estimated to be significant and positive, although with an almost doubling of magnitude. Overall, the LiM SAC model specification yields a clear improvement in goodness of fit to a partial pseudo R^2 of 0.161, in comparison with the SLM specification.

Column (3) then shows the results for a similarity-based social network matrix based on *status homophily*, that is, assigning of weights based on the visible attributes of gender, age, and migrant background. Using this approach, which corresponds to a short-run grouping behavior, entails a minor decrease in the endogenous peer effect parameter to 0.7304, which is statistically significant at the 1% level. The partitioned effects for column (3) show only minor changes as compared to the prior specification. While the effect of using English at home only *sometimes* becomes insignificant, the effect of being born outside of the United States turns significant, although only at the 10% level. Accordingly, the direct effect of being born abroad is estimated to be modest at 4.749 test score points. For the remaining variables, it is striking that the effect of gender on student achievement is reduced in size. While the direct effect decreases by 1.841 points to an 11.845 point achievement gap, the total effect even decreases by almost 9 points in the dependent variable. This specification suggests that the gender achievement gap in mathematics and science can at least partly be explained by within-class grouping behavior.

In Column (4), dyadic similarity is calculated based on *value homophily*, that is, using the students' confidence and interest in mathematics and science, as well as their sense of school belonging as variables. In this specification, which mimicks relatively long-run grouping behavior, the endogenous peer effect parameter increases slightly to 0.7486 and remains significant at the 1% level. The partitioned effects are qualitatively and quantitatively very close to the baseline LiM specification in column (2). Considering *value homophily* rather than the unweighted mean primarily entails a reduction of the gender achievement gap by approximately 1 point. Additionally, this specification provides an overall improvement in fit, as is indicated by the partial pseudo R^2 of 0.164.

Column (5) uses a similarity-based social network matrix that incorporates both the concept of *status* and *value homophily* to emphasize the persistence of initial grouping patterns even in long-run group formation.

[15] In the case of the SAC model, this ratio is based on the endogenous peer effect parameter and therefore fixed.

The endogenous peer effect is estimated at positive 0.7498 and is statistically significant at the 1% level. Again, the partitioned effects exhibit only minor changes compared to both the baseline specification in column (2) as well as the *value homophily* setting in column (4). The direct effect of gender is reduced by 0.830 points compared to the LiM case, thus narrowing the gender achievement gap slightly, although the decrease is not as strong as in the *status homophily* case.

4.2 Adding exogenous peer effects

The remaining columns (6) – (9) alleviate the restriction of individuals being influenced only by peer outcome and not peer attributes. Based on equation (7), this means that the parameter λ is no longer restricted to zero. Instead, peer attributes such as gender, age, frequency of language use, and migrant background now have an effect on individual student test score outcomes. Consequently, indirect effects of variables are allowed to vary in ratio and sign, as compared to the direct effects.

Column (6) presents the LiM calculation of peer averages. In this model specification, the endogenous peer effect is reduced to 0.6930 but remains highly significant. Also, the model increases in terms of goodness of fit, as indicated by the partial pseudo R^2 of 0.175. However, the main changes in the results become apparent when looking at the partitioned effects. The gender achievement gap widens considerably, with a direct effect of 14.563 points, significant at the 1% level. The adverse direct effect of age is reinforced in this specification as well. Most strikingly, the direct effects of language skills show a changing pattern. Having a different mother tongue but *almost always* speaking the language of the test is associated with a positive direct effect amounting to 4.363 points, significant at the 5% level. In contrast, the direct effect of speaking the language only *sometimes* is negative, implying a reduction of scholastic achievement of 9.222 points. The direct effect of perceived bullying remains positive and significant at the 1% level and increases in magnitude. Regarding the indirect effects, it is striking that the negative externality of age is much more pronounced in the nested model (-113.753 points). The externality of bilingual students is again ambivalent. While using the language of test *almost always* poses a positive externality associated with a 58.711 test score point increase, using it only *sometimes* has a negative effect of -117.215 points, both significant at the 1% level. *Never* using the language of the test at home is not associated with a significant effect. However, the indirect effect of being born outside the United States turns significant and positive (66.096 points), although only at the 10% level. Interestingly, the initially puzzling finding of a positive effect of perceived bullying subsists in the nested model, with an even increased magnitude for the indirect effect. Here, the effect of all other students perceiving increased bullying would lead to a 30.121 point increase in the test performance of the individual under study. A possible explanation for this positive externality of perceived bullying is again that high-performing students are more likely to be bullied, yet their performance positively influences their peers.

Peer averages in column (7) are based on the *status homophily* social network matrix. This entails some qualitative and quantitative changes in the partitioned effects, where most strikingly, both the direct and indirect positive effect of using English at home *almost always* becomes insignificant. Additionally, in this specification, a spillover regarding the gender achievement gap can no longer be found. This finding is critical, as this segregating grouping behavior reduces mainly the positive spillovers. If this short-term pattern persists, it bears the risk of creating unfavorable group compositions that are homogeneous in terms of visible characteristics but also low performance.

In contrast, Column (8) shows the results for the social network matrix based on *value homophily*-induced similarity, that is, the long-term group behavior. Relying on this social network matrix increases the goodness of fit to a partial pseudo R^2 of 0.178. While in this specification, the indirect effect of all peers being male regains significance at the 1% level with an effect size of 41.827 points, the adverse effect of not using English regularly is even reinforced. The indirect effects imply 94.599 and 125.493 point reductions for *sometimes* and *never*, although significant only at the 5% and 10% levels, respectively. Apart from that, the direct and indirect effects of speaking English *almost always* become significant and positive again, both at the 10% level. Additionally, the indirect effect of being born outside the United States is reinforced as compared to Column (7). The hypothetical case of all peers being immigrants is associated with a 84.656 point increase in individual test scores, significant at the 5% level. These findings suggest that grouping behavior based on values creates more heterogeneous groups, in which, on the one hand, individuals can benefit from their bilingual peers (as indicated by using English *almost always* and the positive spillover of migratory background), but can, on the other hand, be impeded in their learning by their peers' lower language skills.

The results in column (9) are again based on a combination of *status* and *value homophily*. In this model specification, the results for the partitioned effects are qualitatively very similar to the prior specification. The main difference is observed when considering the indirect effects, where the positive spillover of bilingual students is even increased in magnitude and significance, compared with the *status* and *value homophily* matrices alone. All peers being bilingual and speaking the language of the test *almost always* is associated with a 49.256 test score point increase, significant at the 5% level. In contrast, the adverse externality of lacking language skills (i.e., using English only *sometimes*) amounts to a 110.914-point decrease in the dependent variable. This finding is a weaker form of the pattern found in the LiM model, as becomes visible from the total effects, but it provides a higher goodness of fit, with a partial pseudo R^2 of 0.177.

Overall, relying on the nested model with endogenous and exogenous peer effects yields various important insights. Most strikingly, relaxing the restriction on exogenous peer effects shows the degree to which language skills pose a classroom externality, the magnitude of which depends largely on the peer group composition. Depending on the English language skills of the bilingual peers, this externality can be either positive or negative. This underscores the efficacy of educators' intervention regarding the seating arrangement of students in the classroom and suggests that teachers take language skills into consideration when regrouping students (Freeman et al., 2017). Furthermore, the model specifications suggest that the gender achievement gap could even be more pronounced than in the standard SLM or LiM scenario, although this finding is driven mainly by the indirect effect. To avoid a perpetuation of the gap, intra-class segregation by gender should therefore be avoided.

5 Conclusion

In this paper, I have shown that the application of spatial regression techniques enables an explicit modeling of peer relations in an EPF setting with direct and indirect effects of the respective variables to assess the magnitude of spillovers in a school class as a social network. This strategy furthermore allows for an explicit separate identification of the endogenous and exogenous peer effects described by Manski (1993), as well as an implicit consideration of the correlated factors of the school environment, class characteristics, and idiosyncrasies of the specific mathematics and science teachers. Overall, I find significant and positive endogenous peer effects for eighth grade students in secondary education in the United States. The results are robust to the inclusion of a full set of individual- and parent-level control variables, as well as to fine-grained teacher fixed effects and spatially clustered standard errors at the classroom level.

The peer relations within a class can be approximated by applying theoretical sociology-based approaches of homophily among members of the class as a closed social group. In the absence of information on actual relationships between the nodes of the social network, different measures of similarity can be used to model the probability of interaction among all network members. This focus on more relevant peers within the class strengthens the causal link between individual outcomes and peer group influences and thus increases precision in determining peer groups.

The main findings have proven to be robust to changes in the theoretical foundation of the social network matrix, that is, whether it is based on observable or non-observable characteristics or corresponds to an unweighted averaging of peer attributes. The inclusion of similarity-weighted peer effects helps in reducing omitted variable bias in the estimation of the remaining coefficients and slightly increases the goodness of fit of the overall model. Including peer characteristics as an exogenous peer effect together with the *status* and *value homophily*-based social network matrices gives valuable insights into the mechanisms underlying how, for example, the gender gap in mathematics and science is determined and how linguistic problems and insufficient language preparation affect classroom performance of the individual and also his or her classmates.

The overall results suggest that spatial regression approaches are a suitable choice for estimating peer effects, consistent with recent findings and theoretical approaches (Lee, 2006; Lin, 2010; Boucher et al., 2014). Furthermore, relying on this estimation strategy circumvents the reflection problem discussed at length in the literature on peer effects in education. Describing social networks based on similarity-based relationships among students grouped into classes has proven to be a valid approach to modeling peer relations if no information on reported interactions is available. This increases the range of potential countries that are suitable for studying peer effects and increases the credibility of impacts of peer performance. However, a validation using further country samples and possibly other datasets is left for future research.

References

Aesaert, K. and van Braak, J. (2015). Gender and Socioeconomic Related Differences in Performance Based ICT Competences. *Computers & Education*, 84:8–25.

Anger, S. and Heineck, G. (2010). Do Smart Parents Raise Smart Children? The Intergenerational Transmission of Cognitive Abilities. *Journal of Population Economics*, 23:1255–1282.

Angrist, J. D. (2014). The Perils of Peer Effects. *Labour Economics*, 30:98–108.

Angrist, J. D. and Lang, K. (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, 94(5):1613–1634.

Arcidiacono, P., Foster, G., Goodpaster, N., and Kinsler, J. (2012). Estimating Spillovers Using Panel Data, With An Application to The Classroom. *Quantitative Economics*, 3:421–470.

Badland, H. and Schofield, G. (2006). Understanding the Relationship Between Town Size and Physical Activity Levels: A Population Study. *Health & Place*, 12:538–546.

Bedard, K. and Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *Quarterly Journal of Economics*, 121(4):1437–1472.

Bekiari, A., Pachi, V., and Hasanagas, N. (2017). Investigating Bullying Determinants and Typologies with Social Network Analysis. *Journal of Computer and Communication*, 5:11–27.

Borgna, C. and Contini, D. (2014). Migrant Achievement Penalties in Western Europe: Do Educational Systems Matter? *European Sociologial Review*, 30(5):670–683.

Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2014). Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation. *Journal of Applied Econometrics*, 29:91–109.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of Peer Effects Through Social Networks. *Journal of Econometrics*, 150:41–55.

Brunello, G. and Rocco, L. (2013). The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores. *Economics of Education Review*, 32:234–246.

Burke, M. A. and Sass, T. R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, 31(1):51–82.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3):439–464.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids. *American Economic Journal: Applied Economics*, 2(1):211–228.

Cebolla-Boada, H. (2011). Primary and Secondary Effects in the Explanation of Disadvantage in Education: The Children of Immigrant Families in France. *British Journal of Sociology of Education*, 32(3):407–430.

Cöster, M. E., Fritz, J., Karlsson, C., Rosengren, B. E., and Karlsson, M. K. (2018). Extended Physical Education in Children Aged 6–15 Years Was Associated With Improved Academic Achievement in Boys. *Acta Pædiatrica*, 107:1083–1087.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.

Davis-Keane, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, 19(2):294–304.

Diette, T. M. and Uwaifo Oyelere, R. (2017). Gender and Racial Differences in Peer Effects of Limited English Students: A Story of Language or Ethnicity? *IZA Journal of Migration*, 6(2):1–18.

Fan, X. and Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review*, 13(1):1–22.

Farmer, T. W., Dawes, M., Hamm, J. V., Lee, D., Mehtaji, M., Hoffman, A. S., and Brooks, D. (2018). Classroom Social Dynamics Management: Why the Invisible Hand of the Teacher Matters for Special Education. *Remedial and Special Education*, 39(3):177–192.

Feld, J. F. and Zölitz, U. N. (2016). Understanding Peer Effects - On the Nature, Estimation and Channels of Peer Effects. Working Paper 002, GSBE.

Freeman, R. B., Machin, S., and Viarengo, M. (2010). Variation in Educational Outcomes and Policies Across Countries and of Schools Within Countries. *NBER Working Paper Series*, Working Paper 16293.

Freeman, S., Theobald, R., Crowe, A. J., and Wenderoth, M. P. (2017). Likes Attract: Students self-sort in a Classroom by Gender, Demography, and Academic Characteristics. *Active Learning in Higher Education*, 18(2):115–126.

Fryer, R. G. and Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, 2(2):210–240.

Gaviria, A. and Raphael, S. (2001). School-Based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics*, 83(2):257–268.

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4):857–871.

Grygiel, P., Humenny, G., Rębisz, S., Bajcar, E., and Świtaj, P. (2018). Peer Rejection and Perceived Quality of Relations With Schoolmates Among Children With ADHD. *Journal of Attention Disorders*, 22(8):738–751.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender, and Math. *Science*, 320:1164–1165.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, 18:527–544.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2015). Returns to Skills Around the World: Evidence from PIAAC. *European Economic Review*, 73:103–130.

Hanushek, E. A. and Woessmann, L. (2011). The Economics of International Differences in Educational Achievement. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education. Band 3*, volume 26, pages 89–200. Elsevier.

Harel-Fisch, Y., Walsh, S. D., Fogel-Grinvald, H., Amitai, G., Pickett, W., Molcho, M., Due, P., Gaspar de Matos, M., Craig, W., and Members of the HSBC Violence and Injury Prevention Focus Group (2011). Negative School Perceptions and Involvement in School Bullying: A Universal Relationship Across 40 Countries. *Journal of Adolescence*, 34(4):639–652.

Hughes, J. N., Cao, Q., West, S. G., Smith, P. A., and Cerda, C. (2017). Effect of Retention in Elementary Grades on Dropping Out of School Early. *Journal of School Psychology*, 65:11–27.

Jackson, M., Jonsson, J. O., and Rudolphi, F. (2012). Ethnic Inequality in Choice-Driven Education Systems: A Longitudinal Study of Performance and Choice in England and Sweden. *Sociology of Education*, 85(2):158–178.

Johnsen, A., Bjørknes, R., Iversen, A. C., and Sandbæk, M. (2018). School Competence among Adolescents in Low-Income Families: Does Parenting Style Matter? *Journal of Child and Family Structure*, 27:2285–2294.

Kiss, D. (2017). How Do Ability Peer Effects Operate? Evidence on One Transmission Channel. *Education Economics*, 26(5):1–13.

Knifsend, C. A., Camacho-Thompson, D. E., Juvonen, J., and Graham, S. (2018). Friends in Activities, School-related Affect, and Academic Outcomes in Diverse Middle Schools. *Journal of Youth and Adolescence*, 47:1208–1220.

Kretschmann, J., Vock, M., Lüdtke, O., and Gronostaj, A. (2016). Skipping to the Bigger Pond: Examining Gender Differences in Students' Psychosocial Development after Early Acceleration. *Contemporary Educational Psychology*, 46:195–207.

Kristen, C., Reimer, D., and Kogan, I. (2008). Higher Education Entry of Turkish Immigrant Youth in Germany. *International Journal of Comparative Sociology*, 49(2–3):127–151.

Lee, L.-F. (2006). Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects. *Journal of Econometrics*, 140:333–374.

Lin, X. (2010). Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics*, 28(4):825–860.

Louch, H. (2000). Personal Network Integration: Transitivity and Homophily in Strong-Tie Relations. *Social Networks*, 22:45–64.

Lyle, D. S. (2007). Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *Review of Economics and Statistics*, 89(2):289–299.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542.

Martins, L. and Veiga, P. (2010). Do Inequalities in Parents' Education Play an Important Role in PISA Students' Mathematics Achievement Test Scores Disparities? *Economics of Education Review*, 29:1016–1033.

McCrary, J. and Royer, H. (2011). The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth. *American Economic Review*, 101(1):158–195.

McEwan, P. J. and Shapiro, J. S. (2007). The Benefits of Delayed Primary School Enrollment – Discontinuity Estimates Using Exact Birth Dates. *The Journal of Human Resources*, 43(1):1–28.

McPherson, J. M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.

Moreland, R. L. and Beach, S. R. (1992). Exposure Effects in the Classroom: The Development of Affinity Among Students. *Journal of Experimental Social Psychology*, 28:255–276.

Ponzo, M. (2013). Does Bullying Reduce Educational Achievement? An Evaluation Using Matching Estimators. *Journal of Policy Modeling*, 35:1057–1078.

Resende Oliveira, F., Almeida de Menezes, T., Irffi, G., and Resende Oliveira, G. (2018). Bullying effect on student's performance. *EconomiA*, 19:57–73.

Rindermann, H. (2007). The g-Factor of International Cognitive Ability Comparisons: The Homogeneity of Results in PISA, TIMSS, PIRLS and IQ-Tests Across Nations. *European Journal of Personality*, 21:667–706.

Schuetz, G., Ursprung, H. W., and Woessmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos*, 61:279–308.

Shrum, W., Cheek, N. H., and Hunter, S. M. (1988). Friendship in School: Gender and Racial Homophily. *Sociology of Education*, 61(4):227–239.

Sorensen, L. C., Cook, P. J., and Dodge, K. A. (2017). From Parents to Peers: Trajectories in Sources of Academic Influence Grades 4 to 8. *Educational Evaluation and Policy Analysis*, 39(4):697–711.

Steinberg, L. and Monahan, K. C. (2007). Age Differences in Resistance to Peer Influence. *Dev Psychol*, 43(6):1531–1543.

Todd, P. E. and Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, 113(485):F3–F33.

UNESCO (2019). Global Education Monitoring Report, 2019: Migration, Displacement and Education: Building Bridges, not Walls. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000265866.

Warne, R. T. (2017). Possible Economic Benefits of Full-Grade Acceleration. *Journal of School Psychology*, 65:54–68.

Warne, R. T. and Liu, J. K. (2017). Income Differences Among Grade Skippers and Non-Grade Skippers Across Genders in the Terman Sample, 1936–1976. *Learning and Instruction*, 47:1–12.

Winston, G. and Zimmerman, D. (2004). *Peer Effects in Higher Education*, pages 395–424. University of Chicago Press.

Tables

Table 1: Summary Statistics for Categorical Variables

Variable	Freq.	Percent	Cum.
Student gender			
Female*	4,293	51.19	51.19
Male	4,094	48.81	100.00
Student frequency of using test language at home			
Always*	6,199	73.91	73.91
Almost always	1,419	16.92	90.83
Sometimes	689	8.22	99.05
Never	80	0.95	100.00
Student born in United States			
Yes*	7,932	94.57	94.57
No	455	5.43	100.00
Highest education (Mother)			
Some Primary or Lower secondary*	300	3.58	3.58
Lower secondary	531	6.33	9.91
Upper secondary	1,758	20.96	30.87
Short-cycle tertiary	935	11.15	42.02
Bachelor's or equivalent	1,681	20.04	62.06
Postgraduate degree	1,149	13.70	75.76
Don't know	2,033	24.24	100.00
Highest education (Father)			
Some Primary or Lower secondary*	324	3.86	3.86
Lower secondary	580	6.92	10.78
Upper secondary	1,896	22.61	33.39
Short-cycle tertiary	766	9.13	42.52
Bachelor's or equivalent	1,298	15.48	57.99
Postgraduate degree	946	11.28	69.27
Don't know	2,577	30.73	100.00
Amount of books at home			
0–10 books*	1,429	17.04	17.04
11–25 books	1,820	21.70	38.74
26–100 books	2,477	29.53	68.27
101–200 books	1,396	16.64	84.92
More than 200	1,265	15.08	100.00
Number of digital information devices			
None*	29	0.35	0.35
1–3 devices	426	5.08	5.43
4–6 devices	1,800	21.46	26.89
7–10 devices	2,602	31.02	57.91
More than 10 devices	3,530	42.09	100.00
Number of home study supports			
Neither Own Room nor Internet Connection	148	1.76	1.76
Either Own Room or Internet Connection	1,639	19.54	21.31
Both Own Room and Internet Connection	6,600	78.69	100.00

table continued on next page

Table 1: Summary Statistics for Categorical Variables

Variable	Freq.	Percent	Cum.
Mother born in US			
Yes*	6,434	76.71	76.71
No	1,693	20.19	96.90
I don't know	260	3.10	100.00
Father born in US			
Yes*	6,205	73.98	73.98
No	1,735	20.69	94.67
I don't know	447	5.33	100.00

Notes: The summary statistics above correspond to the unweighted variables from the $8,566$ student sample generated from the TIMSS 2015 release available from the IEA's website. Values that corresponded to missing at random or no answer provided have been removed as is suggested by the corresponding codebook.

* denotes the reference category for categorical variables.

Table 2: Summary Statistics for Metric Variables

Variable	Obs.	Mean	Std.Dev.	Min.	Max.
Mathematics test score	8,387	520.269	78.795	265.531	766.423
Science test score	8,387	531.589	75.942	266.711	754.784
Average test score	8,387	525.929	75.459	273.595	747.156
Student age	8,387	14.235	0.466	9.830	18.330
Perceived bullying	8,387	10.014	1.933	2.467	12.784
Parent home educational resources	8,387	10.826	1.674	4.232	13.884
Student sense of school belonging	8,387	9.594	1.955	3.048	13.622
Student likes mathematics	8,387	9.552	2.060	4.968	13.978
Student confident in mathematics	8,387	10.329	2.330	3.196	15.925
Student likes science	8,387	10.046	2.143	3.771	13.621
Student confident in science	8,387	10.521	2.270	2.821	15.296

Notes: The summary statistics above correspond to the unweighted metric variables from the $8,387$ student sample generated from the TIMSS 2015 release available from the IEA's website. Values that corresponded to missing at random or no answer provided have been removed as is suggested by the corresponding codebook.

Table 3: Pairwise Correlations for the Regression Sample

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Mathematics test score	1.0000										
(2) Science test score	0.9024	1.0000									
(3) Average test score	0.9762	0.9744	1.0000								
(4) Students age	-0.1485	-0.1210	-0.1384	1.0000							
(5) Student bullying	0.0516	0.0606	0.0574	-0.0032	1.0000						
(6) Parents home educational resources	0.4102	0.4357	0.4334	-0.0845	-0.0227	1.0000					
(7) Students sense of school belonging	0.2105	0.1951	0.2081	-0.0277	0.2709	0.1485	1.0000				
(8) Student likes mathematics	0.2771	0.1167	0.2034	-0.0242	0.0590	0.0496	0.3798	1.0000			
(9) Student confident in mathematics	0.4427	0.2900	0.3771	-0.0772	0.1086	0.1575	0.2710	0.6789	1.0000		
(10) Student confident in science	0.1443	0.2730	0.2127	-0.0147	0.0250	0.1501	0.2857	0.1548	0.0529	1.0000	
(11) Student likes science	0.2426	0.3581	0.3069	-0.0325	0.0672	0.2103	0.2132	0.0882	0.1594	0.6929	1.0000

Notes: The pairwise correlations above correspond to the unweighted metric variables from the $8,387$ student sample generated from the TIMSS 2015 release available from the IEA's website.

Table 4: Baseline Regression: Results for Linear-in-Means, Status Homophily, and Value Homophily: Teacher Fixed Effects

	SAC Model					Nested Model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variable	SLM	LiM	SH	VH	SH+VH	LiM	SH	VH	SH+VH
Student gender: Male	12.2406***	12.2094***	10.6677***	11.9137***	11.4456***	12.9850***	12.5430***	12.5396***	11.8225***
	(1.0222)	(1.0304)	(1.0352)	(1.0226)	(1.0246)	(1.0345)	(1.4162)	(1.0250)	(1.0863)
Student age	-11.3951***	-13.1521***	-12.5961***	-12.9888***	-12.8387***	-12.4737***	-10.3346***	-12.3777***	-11.4936***
	(1.1492)	(1.1463)	(1.1533)	(1.1376)	(1.1398)	(1.1576)	(1.2105)	(1.1469)	(1.1591)
Frequency of using language of test at home									
(Reference category: Always)									
Almost always	0.9499	1.6455	1.6727	1.4786	1.5749	2.0420	1.9869	1.7182	2.0071
	(1.6538)	(1.6666)	(1.6735)	(1.6540)	(1.6570)	(1.6745)	(1.6777)	(1.6606)	(1.6626)
Sometimes	-3.6568	-3.9544*	-3.8055	-4.0652*	-3.9671*	-4.5884*	-4.2468*	-4.5215*	-4.5276*
	(2.3175)	(2.3520)	(2.3596)	(2.3338)	(2.3383)	(2.3910)	(2.3846)	(2.3651)	(2.3676)
Never	-1.4086	-5.5530	-4.6318	-5.5006	-5.3300	-1.6123	-1.7045	-1.7894	-1.3901
	(5.4538)	(5.5668)	(5.5867)	(5.5243)	(5.5346)	(5.5918)	(5.6020)	(5.5441)	(5.5518)
Student Born Outside of US	2.0231	3.7700	4.2765*	3.5236	3.7722	2.4454	2.6995	2.4972	2.3821
	(2.4627)	(2.4908)	(2.5003)	(2.4716)	(2.4762)	(2.5062)	(2.6070)	(2.4859)	(2.4947)
Perceived Bullying	0.6703**	1.1527***	1.1243***	1.0570***	1.0890***	1.0183***	1.0270***	0.8858***	0.9293***
	(0.2655)	(0.2651)	(0.2665)	(0.2631)	(0.2636)	(0.2674)	(0.2680)	(0.2651)	(0.2656)
Endogenous Peer Effect		0.7476***	0.7304***	0.7486***	0.7498***	0.6930***	0.6872***	0.6949***	0.6986***
		(0.0102)	(0.0115)	(0.0090)	(0.0096)	(0.0136)	(0.0140)	(0.0124)	(0.0128)
Spatial Error	0.3836***	-1.6152***	-1.4602***	-1.5842***	-1.5994***	-1.5553***	-1.4125***	-1.4861***	-1.5274***
	(0.0204)	(0.0627)	(0.0762)	(0.0630)	(0.0634)	(0.0716)	(0.0800)	(0.0736)	(0.0725)
Peer gender: Male						3.7533	-4.7153	4.5714	4.4952
						(4.7694)	(4.4190)	(4.3727)	(4.5283)
Peer age						-27.6936***	-24.9999***	-23.9458***	-26.8300***
						(4.6193)	(4.4932)	(4.3202)	(4.4951)
Peer frequency of using language of test at home									
(Reference category: Always)									
Almost always						17.3395**	6.8417	12.2201*	14.0525**
						(7.2728)	(6.4524)	(6.7309)	(6.7803)
Sometimes						-34.2636**	-27.3591**	-26.9039**	-31.6377**
						(14.1725)	(11.2205)	(12.2444)	(12.4573)
Never						-33.2101	-28.7914	-38.6148*	-26.0750
						(25.4409)	(18.3809)	(22.7381)	(20.7519)
Peer Born Outside of US						19.4191*	15.6420*	25.1527**	16.7694
						(11.5435)	(8.6905)	(10.7914)	(10.9006)
Parental Factors?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,387	8,387	8,387	8,387	8,387	8,387	8,387	8,387	8,387
Log Likelihood	-44006	-43911	-43930	-43850	-43868	-43840	-43873	-43780	-43797
Pseudo R2	0.106	0.161	0.155	0.164	0.163	0.175	0.168	0.178	0.177

Notes: Dependent variable is the average score in the TIMSS subjects of mathematics and science. Constant term is included, but not shown.
Parental Factors refer to controls for educational achievement, immigration background, and home economic and educational resources specific to the student's parents.
Teacher Fixed Effects refer to indicator variables for each assigned teacher ID which are absorbed in an auxiliary regression. Note that classes can have more than one teacher per subject.

Robust standard errors, clustered by classroom using the adjacency social network matrix, are reported in parentheses.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 5: Baseline Regression: Overview of Direct, Indirect, and Total Impacts

	SAC Model					Nested Model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variable	SLM	LiM	SH	VH	SH+VH	LiM	SH	VH	SH+VH
	Direct Effects								
Student gender: Male	12.241***	13.686***	11.845***	13.368***	12.856***	14.563***	13.025***	14.206***	13.451***
	(1.022)	(1.158)	(1.15)	(1.15)	(1.153)	(1.286)	(1.233)	(1.253)	(1.135)
Student age	-11.395***	-14.742***	-13.987***	-14.575***	-14.421***	-16.971***	-14.295***	-16.459***	-15.945***
	(1.149)	(1.283)	(1.279)	(1.275)	(1.279)	(1.322)	(1.262)	(1.316)	(1.287)
Frequency of using language of test at home									
(Reference category: Always)									
Almost always	0.950	1.845	1.857	1.659	1.769	4.363**	2.999	3.400*	3.981*
	(1.654)	(1.868)	(1.858)	(1.856)	(1.861)	(2.071)	(2.021)	(2.036)	(2.044)
Sometimes	-3.657	-4.433*	-4.226	-4.562*	-4.456*	-9.222***	-7.982***	-8.289***	-8.971***
	(2.317)	(2.636)	(2.62)	(2.619)	(2.626)	(3.351)	(3.091)	(3.185)	(3.215)
Never	-1.409	-6.225	-5.143	-6.172	-5.987	-5.861	-5.401	-6.788	-4.844
	(5.454)	(6.24)	(6.203)	(6.199)	(6.217)	(6.998)	(6.576)	(6.796)	(6.696)
Student Born Outside of US	2.023	4.226	4.749*	3.954	4.237	5.058	4.858*	5.869*	4.736
	(2.463)	(2.791)	(2.776)	(2.773)	(2.781)	(3.138)	(2.776)	(3.087)	(2.961)
Perceived Bullying	0.670**	1.292***	1.248***	1.186***	1.223***	2.209***	2.064***	2.067***	2.171***
	(0.265)	(0.297)	(0.296)	(0.295)	(0.296)	(0.311)	(0.312)	(0.307)	(0.309)
	Indirect Effects								
Student gender: Male		34.652***	27.693***	33.979***	32.867***	39.918***	11.99	41.827***	40.649***
		(3.44)	(3.083)	(3.315)	(3.357)	(15.09)	(11.756)	(13.955)	(13.759)
Student age		-37.328***	-32.699***	-37.045***	-36.867***	-113.753***	-98.582***	-102.473***	-111.101***
		(3.716)	(3.448)	(3.613)	(3.684)	(13.203)	(12.2)	(12.769)	(13.007)
Frequency of using language of test at home									
(Reference category: Always)									
Almost always		4.67	4.342	4.217	4.522	58.711**	25.204	42.234*	49.256**
		(4.74)	(4.354)	(4.724)	(4.766)	(23.220)	(20.306)	(21.710)	(22.098)
Sometimes		-11.223*	-9.879	-11.594*	-11.392*	-117.215***	-92.980***	-94.599**	-110.914***
		(6.693)	(6.144)	(6.672)	(6.731)	(44.856)	(35.100)	(39.332)	(40.34)
Never		-15.760	-12.024	-15.688	-15.305	-107.461	-92.013	-125.493*	-86.199
		(15.816)	(14.516)	(15.768)	(15.907)	(81.185)	(57.962)	(73.016)	(67.614)
Student Born Outside of US		10.700	11.101*	10.049	10.832	66.096*	53.732**	84.656**	58.750*
		(7.076)	(6.507)	(7.055)	(7.119)	(36.443)	(25.697)	(34.350)	(34.521)
Perceived Bullying		3.272***	2.919***	3.015***	3.127***	30.121***	25.810***	29.659***	30.991***
		(0.767)	(0.707)	(0.760)	(0.769)	(3.287)	(3.139)	(3.168)	(3.248)
	Total Effects								
Student gender: Male	12.241***	48.338***	39.538***	47.347***	45.723***	54.481***	25.014**	56.033***	54.100***
	(1.022)	(4.498)	(4.141)	(4.386)	(4.426)	(15.911)	(11.541)	(14.734)	(14.173)
Student age	-11.395***	-52.070***	-46.686***	-51.620***	-51.288***	-130.724***	-112.877***	-118.932***	-127.046***
	(1.149)	(4.890)	(4.612)	(4.801)	(4.865)	(13.909)	(12.622)	(13.486)	(13.617)
Frequency of using language of test at home									
(Reference category: Always)									
Almost always	0.9500	6.515	6.200	5.876	6.291	63.074**	28.203	45.634**	53.236**
	(1.654)	(6.607)	(6.211)	(6.579)	(6.627)	(24.544)	(21.554)	(23.001)	(23.396)
Sometimes	-3.657	-15.656*	-14.104	-16.156*	-15.848*	-126.437***	-100.962***	-102.888**	-119.885***
	(2.317)	(9.323)	(8.758)	(9.285)	(9.352)	(47.356)	(37.249)	(41.632)	(42.676)
Never	-1.409	-21.985	-17.167	-21.860	-21.292	-113.321	-97.414	-132.281*	-91.043
	(5.454)	(22.051)	(20.716)	(21.963)	(22.120)	(85.722)	(61.913)	(77.317)	(71.758)
Student Born Outside of US	2.023	14.926	15.850*	14.003	15.069	71.154*	58.590**	90.526**	63.486*
	(2.463)	(9.863)	(9.276)	(9.824)	(9.895)	(38.481)	(26.781)	(36.343)	(36.218)
Perceived Bullying	0.6700**	4.564***	4.167***	4.201***	4.350***	32.330***	27.874***	31.726***	33.162***
	(0.2650)	(1.060)	(0.999)	(1.052)	(1.061)	(3.462)	(3.316)	(3.337)	(3.420)

Notes: Partitioned effects are calculated based on the estimated regression coefficients presented in table 4 (see section B.1 of the appendix for details). The dependent variable is the average score in the TIMSS subjects of mathematics and science.
Robust standard errors are reported in parentheses.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

A Figures

Figure 1: Example of similarities within a class

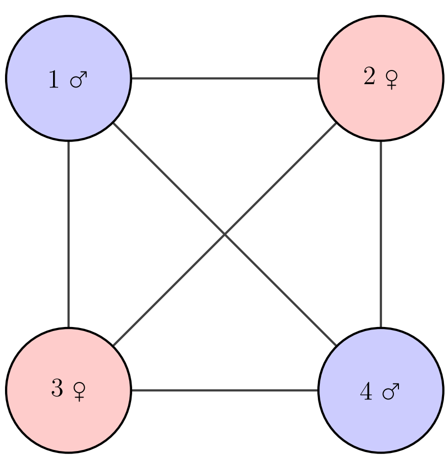
(a) Student characteristics

Student i	Gender	Age	Attitude
1	male	14.1	2.7
2	female	14.7	3.2
3	female	14.0	9.1
4	male	14.5	5.3

(b) Dyadic Gower similarity measures

		Student 1		*Student 2*		*Student 3*
Student 2	sex	0.00				
	age	0.14				
	attitude	0.92				
	total	0.35				
Student 3	sex	0.00	sex	1.00		
	age	0.86	age	0.00		
	attitude	0.29	attitude	0.08		
	total	0.38	**total**	0.36		
Student 4	sex	1.00	sex	0.00	sex	0.00
	age	0.43	age	0.71	age	0.29
	attitude	0.59	attitude	0.67	attitude	0.41
	total	0.67	**total**	0.46	**total**	0.23

(c) Network visualization: Linear-in-Means



(d) Network visualization: Similarity-based

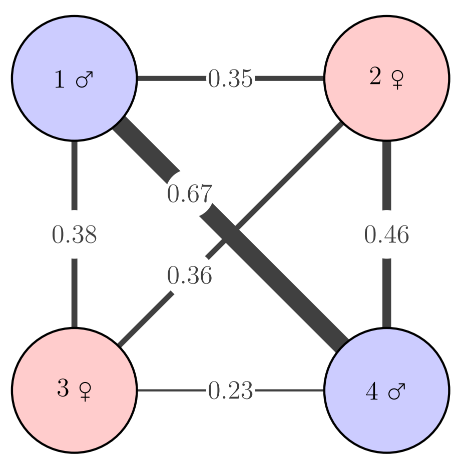
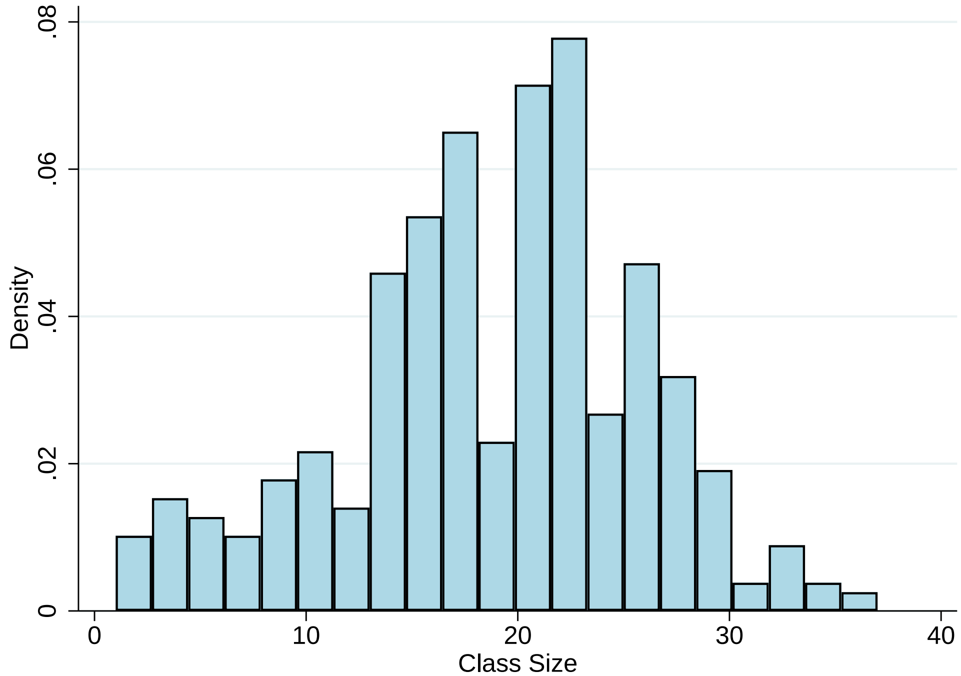


Figure 2: Histogram of Class Sizes in the Estimation Sample



B Technical appendix

B.1 Interpreting parameter estimates

In spatial regression models, coefficient estimates cannot be simply interpreted as marginal effects due to spillover or feedback mechanisms. Rather, effects of changes in the explanatory variables can be partitioned into direct, indirect, and total effects, which differ from the estimated regression coefficients. The derivation of the partitioned effects is illustrated using an example based on the spatial autoregressive model specification. For an SAR model, a simplified version of the regression equation (7), solved for O, can be portrayed as equation (8), where $\mathbf{Z}$ is a matrix of all K variables[16] used in the prior regression and θ is the corresponding coefficient vector.

$$O = (\mathbf{I_n} - \rho\mathbf{W})^{-1}\left[\alpha\iota_{\mathbf{n}} + \mathbf{Z}\theta + \varepsilon\right] \tag{8}$$

Refraining from the matrix notation for $\mathbf{Z}\theta$, the equation can be rearranged and rewritten so that z_r is now a $n \times 1$ vector for the observations of variable $r \in [1, \ldots, K]$.

$$O = V(\mathbf{W})\left[\alpha\iota_{\mathbf{n}} + \varepsilon\right] + \sum_{r=1}^{K} S_r(\mathbf{W})z_r \tag{9}$$

$$\text{with } S_r(\mathbf{W}) = V(\mathbf{W})(\mathbf{I_n}\theta_r)$$

$$\text{and } V(\mathbf{W}) = (\mathbf{I_n} - \rho\mathbf{W})^{-1}$$

Expanding the matrices $S_r(\mathbf{W})$, $\mathbf{O}$ and vector z_r yields the following equation.

$$\begin{pmatrix} o_1 \\ o_2 \\ \vdots \\ o_n \end{pmatrix} = \sum_{r=1}^{K} \begin{bmatrix} S_r(\mathbf{W}_{11}) & S_r(\mathbf{W}_{12}) & \cdots & S_r(\mathbf{W}_{1n}) \\ S_r(\mathbf{W}_{21}) & S_r(\mathbf{W}_{22}) & \cdots & S_r(\mathbf{W}_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ S_r(\mathbf{W}_{n1}) & S_r(\mathbf{W}_{n2}) & \cdots & S_r(\mathbf{W}_{nn}) \end{bmatrix} \begin{pmatrix} z_{1r} \\ z_{2r} \\ \vdots \\ z_{nr} \end{pmatrix} + V(\mathbf{W})\left[\alpha\iota_{\mathbf{n}} + \mathbf{WX}\lambda + \varepsilon\right] \tag{10}$$

For a single observation, for example, student i, this can be rewritten as

$$o_i = \sum_{r=1}^{K} \left[S_r(\mathbf{W})_{i1}z_{1r} + S_r(\mathbf{W})_{i2}z_{2r} + \cdots + S_r(\mathbf{W})_{in}z_{nr}\right] + V(\mathbf{W})_i\left[\alpha + \varepsilon\right] \tag{11}$$

Taking the derivative with respect to variable r of individual j then yields the following expression

$$\frac{\partial o_i}{\partial z_{jr}} = S_r(\mathbf{W})_{ij} \tag{12}$$

[16] This includes student and parent-level variables

For variable r, the effect of a change of individual $j \in [1, \ldots, n]$ on individual $i \in [1, \ldots, n]$ can be portrayed in a $n \times n$ matrix of derivatives. For $i \neq j$, this matrix illustrates that, in contrast to the standard OLS case, the derivative of o_i with respect to z_{jr} is non-zero. Changes in characteristics of individual j affect not only j's outcome but also the outcomes of all other individuals.

$$S_r(\mathbf{W}) = \begin{bmatrix} \frac{\partial o_1}{\partial z_{1r}} & \frac{\partial o_1}{\partial z_{2r}} & \cdots & \frac{\partial o_1}{\partial z_{nr}} \\ \frac{\partial o_2}{\partial z_{1r}} & \frac{\partial o_2}{\partial z_{2r}} & \cdots & \frac{\partial o_2}{\partial z_{nr}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial o_n}{\partial z_{1r}} & \frac{\partial o_n}{\partial z_{2r}} & \cdots & \frac{\partial o_n}{\partial z_{nr}} \end{bmatrix} = \begin{bmatrix} S_r(\mathbf{W})_{11} & S_r(\mathbf{W})_{12} & \cdots & S_r(\mathbf{W})_{1n} \\ S_r(\mathbf{W})_{21} & S_r(\mathbf{W})_{22} & \cdots & S_r(\mathbf{W})_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_r(\mathbf{W})_{n1} & S_r(\mathbf{W})_{n2} & \cdots & S_r(\mathbf{W})_{nn} \end{bmatrix} \tag{13}$$

From the matrix in equation (13), the partitioned effects can be read in the following way. First, the *direct effects* can be found on the main diagonal of the matrix. The average direct effect corresponds to the average of all elements on the main diagonal, that is,

$$\overline{direct}_r = n^{-1} \operatorname{tr}\left(S_r(\mathbf{W})\right) \tag{14}$$

The *total effect* on individual i resulting from a one-unit change in variable z_r for all individuals (including i) can be read from the sum of the ith row of the matrix. Accordingly, the average total effect corresponds to the average of the sum of all rows, which can be expressed as in equation (15).

$$\overline{total}_r = n^{-1} \iota'_n \operatorname{tr}\left(S_r(\mathbf{W})\right) \iota_n \tag{15}$$

Last, the *indirect effect* on the outcome of student i corresponds to a change in variable z_r for all other individuals, except for i. This corresponds to the difference between the individual total and direct effect. The average indirect effect then is the average total minus the average direct effect.

$$\overline{indirect}_r = \overline{total}_r - \overline{direct}_r \tag{16}$$

The results presented in table 5 have been constructed according to the formulas introduced above, using the built-in post-estimation command *estat impact* from the *sp* class commands introduced in Stata version 15. By default, this command uses the delta method to calculate the variance of the impacts.

Zuletzt erschienen /previous publications:

V-428-20 **Bernhard C. Dannemann,** Peer Effects in Secondary Education: Evidence from the 2015 Trends in Mathematics and Science Study Based on Homophily

V-427-19 **Christoph Böhringer, Knut Einar Rosendahl, Halvor Briseid Storrøsten,** Smart hedging against carbon leakage

V-426-19 **Emmanuel Asane-Otoo, Bernhard Dannemann,** Rockets and Feathers Revisited: Asymmetric Retail Fuel Pricing in the Era of Market Transparency

V-425-19 **Heinz Welsch,** Moral Foundations and Voluntary Public Good Provison: The Case of Climate Change

V-424-19 **Gökçe Akın-Olçum, Christoph Böhringer, Thomas Rutherford, Andrew Schreiber,** Economic and Environmental Impacts of a Carbon Adder in New York

V-423-19 **Jasper N. Meya, Paul Neetzow,** Renewable energy policies in federal government systems

V-422-19 **Philipp Biermann, Heinz Welsch**, Changing Conditions, Persistent Mentality: An Anatomy of East German Unhappiness, 1990-2016

V-421-19 **Philipp Biermann, Jürgen Bitzer, Erkan Gören,** The Relationship between Age and Subjective Well-Being: Estimating Within and Between Effects Simultaneously

V-420-19 **Philipp Poppitz,** Multidimensional Inequality and Divergence: The Eurozone Crisis in Retrospect

V-419-19 **Heinz Welsch,** Utilitarian and Ideological Determinants of Attitudes toward Immigration: Germany before and after the “Refugee Crisis”

V-418-19 **Christoph Böhringer**, **Xaquin Garcia-Muros, Mikel González-Eguino**, Greener and Fairer: A Progressive Environmental Tax Reform for Spain

V-417-19 **Heinz Welsch, Martin Binder, Ann-Kathrin Blankenberg**, Pro-environmental norms and subjective well-being: panel evidence from the UK

V-416-18 **Jasper N. Meya,** Environmental Inequality and Economic Valuation

V-415-18 **Christoph Böhringer, Thomas F. Rutherford, Edward J. Balistreri,** Quantifying Disruptive Trade Policies

V-414-18 **Oliver Richters, Andreas Siemoneit**, The contested concept of growth imperatives: Technology and the fear of stagnation

V-413-18 **Carsten Helm, Mathias Mier**, Subsidising Renewables but Taxing Storage? Second-Best Policies with Imperfect Carbon Pricing

V-412-18 **Mathias Mier**, Policy Implications of a World with Renewables,Limited Dispatchability, and Fixed Load

V-411-18 **Klaus Eisenack, Mathias Mier**, Peak-load Pricing with Different Types of Dispatchability

V-410-18 **Christoph Böhringer, Nicholas Rivers,** The energy efficiency rebound effect in general equilibrium

V-409-18 **Oliver Richters, Erhard Glötzl**, Modeling economic forces, power relations, and stock-flow consistency: a general constrained dynamics approach

V-408-18 **Bernhard C. Dannemann, Erkan Gören,** The Educational Burden of ADHD: Evidence From Student Achievement Test Scores

V-407-18 **Jürgen Bitzer, Erkan Gören**, Foreign Aid and Subnational Development: A Grid Cell Analysis

V-406-17 **Christoph Böhringer**, **Jan Schneider**, **Marco Springmann,** Economic and Environmental Impacts of Raising Revenues for Climate Finance from Public Sources

V-405-17 **Erhard Glötzl, Florentin Glötzl, Oliver Richters**, From constrained optimization to constrained dynamics: extending analogies between economics and mechanics

V-404-17 **Heinz, Welsch, Jan Kühling**, How Green Self Image Affects Subjective Well-Being: Pro-Environmental Values as a Social Norm

V-403-17 **Achim Hagen, Jan Schneider**, Boon or Bane? Trade Sanctions and the Stability of International Environmental Agreements

V-402-17 **Erkan Gören**, The Role of Novelty-Seeking Traits in Contemporary Knowledge Creation
V-401-17 **Heinz Welsch, Jan Kühling**, Divided We Stand: Immigration Attitudes, Identity, and Subjective Well-Being
V-400-17 **Christoph Böhringer, Thomas F. Rutherford**, Paris after Trump: An inconvenient insight
V-399-17 **Frank Pothen, Heinz Welsch**, Economic Development and Material Use
V-398-17 **Klaus Eisenack, Marius Paschen,** Designing long-lived investments under uncertain and ongoing change
V-397-16 **Marius Paschen**, The effect of intermittent renewable supply on the forward premium in German electricity markets
V-396-16 **Heinz Welsch, Philipp Biermann**, Poverty is a Public Bad: Panel Evidence from Subjective Well-being Data
V-395-16 **Philipp Biermann**, How Fuel Poverty Affects Subjective Well-Being: Panel Evidence from Germany
V-394-16 **Heinz Welsch**, Electricity Externalities, Siting, and the Energy Mix: A Survey
V-393-16 **Leonhard Kähler, Klaus Eisenack**, Strategic Complements in International Environmental Agreements: a New Island of Stability
V-392-16 **Christoph Böhringer, Xaquin Garcia-Muros, Ignacio Cazcarro, Iñaki Arto**, The Efficiency Cost of Protective Measures in Climate Policy
V-391-16 **Achim Hagen, Juan-Carlos Altamirano-Cabrera, Hans-Peter Weikard**, The Influence of Political Pressure Groups on the Stability of International Environmental Agreements
V-390-16 **Christoph Böhringer, Florian Landis, Miguel Angel Tovar Reaños,** Cost-effectiveness and Incidence of Renewable Energy Promotion in Germany
V-389-16 **Carsten Helm, Mathias Mier,** Efficient diffusion of renewable energies: A roller-coaster ride
V-388-16 **Christoph Böhringer, Jan Schneider, Emmanuel Asane-Otoo,** Trade In Carbon and The Effectiveness of Carbon Tariffs
V-387-16 **Achim Hagen, Leonhard Kähler, Klaus Eisenack,** Transnational Environmental Agreements with Heterogeneous Actors
V-386-15 **Jürgen Bitzer, Erkan Gören, Sanne Hiller,** Absorption of Foreign Knowledge: Firms' Benefits of Employing Immigrants
V-385-15 **Klaus Eisenack, Julien Minnemann, Paul Neetzow, Felix Reutter,** Contributions to the institutional economics of the energy transition
V-384-15 **Christoph Böhringer, Xaquín Garcia-Muros, Mikel Gonzalez-Eguino, Luis Rey,** US Climate Policy: A Critical Assessment of Intensity Standards
V-383-15 **Christoph Böhringer, Edward J. Balistreri, Thomas F. Rutherford**, Carbon policy and the structure of global trade
V-382-15 **Christoph Böhringer, Brita Bye, Taran Fæhn, Knut Einar Rosendahl,** Output-based rebating of carbon taxes in the neighbor's backyard
V-381-15 **Christoph Böhringer, Markus Bortolamedi**, Sense and No(n)-Sense of Energy Security Indicators
V-380-15 **Christoph Böhringer, Knut Einar Rosendahl, Halvor Briseid Storrøsten,** Mitigating carbon leakage: Combining output-based rebating with a consumption tax
V-379-15 **Jan Micha Steinhäuser, Klaus Eisenack**, Spatial incidence of large-scale power plant curtailment costs
V-378-15 **Carsten Helm, Franz Wirl**, Climate policies with private information: The case for unilateral action
V-377-15 **Klaus Eisenack**, Institutional adaptation to cooling water scarcity in the electricity sector under global warming
V-376-15 **Christoph Böhringer, Brita Bye, Taran Fæhn, and Knut Einar Rosendahl**, Targeted carbon tariffs – Carbon leakage and welfare effects
V-375-15 **Heinz Welsch, Philipp Biermann**, Measuring Nuclear Power Plant Externalities Using Life Satisfaction Data: A Spatial Analysis for Switzerland
V-374-15 **Erkan Gören**, The Relationship Between Novelty-Seeking Traits And Comparative Economic Development