Exploring Mechanisms of Peer Effects in Education
– Frame Factor Analyses of Classroom Instruction

Pontus Bäckström
Exploring Mechanisms of Peer Effects in Education
– Frame Factor Analyses of Classroom Instruction

Pontus Bäckström
The classroom is a place of high drama. You’ll never know what you’ve done to, or for, the hundreds coming and going. You see them leaving the classroom: dreamy, flat, sneering, admiring, smiling, puzzled. After a few years you develop antennae. You can tell when you’ve reached them or alienated them. It’s chemistry. It’s psychology. It’s animal instinct.

Frank McCourt, *Teacher Man*
Abstract

Despite a vast international literature, previous research into compositional effects and peer effects in education has remarked on the lack of consensus within the field. To date, peer effects in education and students’ different opportunities to learn in school have been studied separately. In this thesis, it is argued that these perspectives need to be synthesized.

In the thesis, Frame Factor Theory is employed as a theory of peer effects in classroom instruction. According to the theory, one mechanism generating peer effects is the steering and limiting effect that class composition has on teachers’ instruction. It is argued that this perspective also needs to be synthesized with the Opportunity to Learn-perspective, to widen our understanding of why different students meet different instruction and thus are given different opportunities to learn. The theoretical perspectives and empirical models are evolved through three empirical studies included in the thesis. The theoretical development and the empirical results from each study is discussed in an integrative essay in the thesis.

The results indicate that class composition affects the presence of limitations on instruction and content coverage in instruction, both of which is related to students’ opportunities to learn and individual students’ achievement. Implications for future research and educational policy is discussed.
Original papers

The following papers are enclosed as appendices.

Paper 1

School Composition, Disruptive Classroom Behaviour and Student Results: A Study of Mechanisms of Peer Effects

Pontus Bäckström

Paper 2

Exploring Mechanisms of Peer Effects in Education: A Frame-Factor Analysis of Instruction

Pontus Bäckström

Paper 3


Pontus Bäckström
Table of Contents

Preface ................................................................................................................................. 1

1. **Introduction** ............................................................................................................... 4
   1.1. The rationale of the thesis ......................................................................................... 5
   1.2. Key concepts ............................................................................................................. 6
   1.3. Research aims .......................................................................................................... 7
   1.4. Disposition of the thesis ......................................................................................... 9

2. **Background – The Swedish setting** ............................................................................. 11
   2.1. Educational reform in Sweden .............................................................................. 11
   2.2. Achievement, equity and peer effects in education .............................................. 14

3. **Previous research** ......................................................................................................... 16
   3.1. Compositional effects and peer effects in education ............................................ 16
   3.2. Models of compositional effects ............................................................................. 19
       3.2.1. The school compositional hypothesis ................................................................. 19
       3.2.2. Conceptual model of peer influences on learning ........................................... 20
       3.2.3. The Opdenakker model of compositional effects ............................................ 22
       3.2.4. TIMSS model of potential educational experiences ....................................... 23
   3.3. Towards a theory of peer effects ............................................................................ 25

4. **Theoretical framework** .................................................................................................. 26
   4.1. Frame Factor Theory .............................................................................................. 26
   4.2. Opportunity to learn .............................................................................................. 32
   4.3. The need for theoretical synthesis ......................................................................... 37

5. **Methods and data** ......................................................................................................... 40
   5.1. Study I .................................................................................................................... 40
       5.1.1. Data .................................................................................................................. 41
5.1.2. Analytical strategy and methods ......................................................... 41
5.2. Studies II and III .............................................................................. 41
  5.2.1. Structural equation modelling ....................................................... 42
  5.2.2. Assessing models within SEM ....................................................... 43
  5.2.3. Data .............................................................................................. 46
  5.2.4. Analytical strategy and methods ....................................................... 48
5.3. Validity and reliability .......................................................................... 49
  5.3.1. Establishing validity and reliability ................................................... 49
  5.3.2. Validity and reliability of Swedish school administrative data 50
  5.3.3. Validity and reliability of TIMSS ....................................................... 51
  5.3.4. Concluding remarks on validity and reliability ............................. 57
5.4. Research ethics .................................................................................... 58

6. Empirical findings and integrated discussion of original studies ................................................................................................................. 60
  6.1. Study I: School Composition, Disruptive Classroom Behaviour and Student Results: A Study of Mechanisms of Peer Effects ......................................................... 61
  6.4. Integrated discussion ......................................................................... 65
    6.4.1. Summarizing and interpreting the empirical findings .............. 65
    6.4.2. Future research ........................................................................ 68
    6.4.3. In closure .................................................................................. 69

References ........................................................................................................ 73

Original papers
Preface

It is January. The car is packed, the lunch is prepared, and the family is, after some trouble, seated in the car. We are going to the hockey arena for some ice-skating during the public hours. The children often enjoy skating, at least when we are there. There is always a need for some initial persuasion, and that is why it is a relief that we are on the way.

After arrival, when we are in the midst of the sweat dripping exercise of tying all skates, we realize that we have forgotten the children’s coveralls. Me and my wife exchange a swift glance; what do we do now? It is after all a few degrees below the freezing point in the arena and the children are only dressed in their first layer underwear. Luckily, the family have for a long time, more or less as a joke, adopted the US Marine Corps unofficial motto of “improvise, adapt and overcome”, often also accompanied with the war cry “Oorah!”. We search our bags frantically, trying to find a solution. The solution turns out to be that one of the kids must skate in my hunting parkas, brought along to keep us warm during the lunch break, and the other one must skate in me and my wife’s second layer fleeces. It looks so funny. The sleeves are so long that their hands are well-hidden beneath all fabric, and their legs are covered beneath their knees. They look like two tadpoles as they happily through themselves out on the ice with a loud war cry, Oorah!

The sense morale of this little anecdote relates to the main theme of this thesis. No matter how well-planned classroom instruction might be, things always happen which requires the teacher to be able to improvise, adopt and overcome. In Frank McCourt’s words, “the classroom is a place of high drama”. It is a dynamic jumble of correlations and causality in which the students and the teacher all affect each other in different ways.

Given this, one main message of this thesis is that we, within classroom instruction, must understand that the teacher’s instruction to some extent is dependent on the conditions given. The instruction is framed by its’ conditions. In our analyses of classroom instruction, we cannot interpret the teacher as an independent variable – a robot – who mechanically can
implement the planned lesson and reach all aims postulated. One of the main conditions affecting teachers’ instruction is the students’ prior knowledge. In many cases, the students’ prior knowledge is a necessary condition for their forthcoming learning; how do you teach a student to calculate percentages if the student do not know how to calculate multiplication or division? The main argument of the thesis is that these circumstances are one mechanism generating peer effects within classroom instruction on individual students’ achievement.

With this thesis I conclude my third-cycle education. It has been a great privilege to be able to do this at this point in my life, in the middle of everything. This would not have been possible if it were not for some special people.

First of all, it would not have been possible if it were not for my boss Andreas Mörck, general secretary for The Swedish Teachers’ Union, who has always encouraged me, had faith in me, and always been able to give me the right conditions to be able to complete this project. My warm and sincere thanks to you!

Second, it would not have been possible if it were not for my supervisors Jesper Boesen and Charlotta Mellander. You have completed each other in so many ways! Jesper, you have always made sure that we look beyond the horizon, that we have seen the full picture and that we have taken the next step. Charlotta, you have always made necessary halts, reviewed what I was doing and always made me question if other people would understand what I am trying to do in this thesis. My warm and sincere thanks to both of you!

But above all, and most important, was and always will be my family. Mirabelle and Molly, nothing compares to the joy and pride of being your father. I think that we have managed to balance your homework in mathematics, your studies for history exams, my third-cycle education and our mutual goal to eat all dinners together as a family with success.

And Lina, my life partner, you have been obliged to take numerous walks alone in the dark with the dogs, for which I will always be in dept to you. But we have also been able to find time to go and sleep in the woods, have champagne in our croft and paddle canoe. Life with you is always a joy. Our
first 25 years together have been a blast, and nothing makes me happier than the thought of us having more than 25 years left together!

Pontus Bäckström

Tumba, January 2024
1. Introduction

In Frank McCourt’s memoir ‘Teacher Man’ (2005), there is a scene where McCourt speaks to a new colleague about what it is like to work as a teacher. McCourt says:

The classroom is a place of high drama. You’ll never know what you’ve done to, or for, the hundreds coming and going. You see them leaving the classroom: dreamy, flat, sneering, admiring, smiling, puzzled. After a few years, you develop antennae. You can tell when you’ve reached them or alienated them. It’s chemistry. It’s psychology. It’s animal instinct. (p. 255)

This quote reflects much of the problems educational research is facing. What happens within the classrooms – the ‘high drama’ – is a complex, dynamic compote of intra- and interpersonal correlations and causality, a puzzle almost impossible to solve. This is a reason why some researchers have called it a ‘black box’ problem (Wilkinson & Fung, 2002; Lavy, Paserman, & Schlosser, 2011; Liu, Van Damme, Gielen, & Van Den Noortgate, 2015). We know it’s there, we know there’s an output, but we’re not exactly sure how and why things turn out as they do.

Of course, we’re not groping in complete darkness. We have several clues concerning how and why some students learn (or don’t learn) the intended curricula. We know that it makes a difference what background the students have (Sirin, 2005), what teacher they have (Johansson, Gustafsson, Hansson, & Alatalo, 2023), what their teacher does in instruction and so on (Hattie, 2009).

We also know that it is important to individual students which class they attend, or more precisely, who their classmates are (Sacerdote, 2011). Research in this field is commonly labelled ‘peer effects’ or ‘compositional effects’; the usage differs between different research traditions (Dreeben & Barr, 1988; Thrupp, Lauder, & Robinson, 2002; Epple & Romano, 2011; Televantou, et al., 2015). Yet, the issue at the heart of the research remains the
same: How important are classmates for an individual student’s achievement? (Arnott & Rowse, 1987)

Despite the vast literature on peer effects and compositional effects in education, several researchers have noticed the lack of consensus within this field. Thrupp, Lauder & Robinson (2002) note that ‘there is remarkably little consensus over the nature and size of school compositional and peer effects’ (abstract). A similar note is made by Rutter & Maughan (2002), who claim that

Despite a substantial growth in studies of peer influences in the behavioural literature, there has been much less investigation of the ways in which they operate in schools or of their effects on school functioning or pupil progress. (p. 469)

One reason for this lack of consensus, according to Thrupp, Lauder & Robinson (2002), is that school compositional effects have been ‘inadequately theorised and poorly operationalised’ (p. 484).

These problems aside, it is a well-known fact that students perform differently in school. It is equally well known that it matters what which school a student attends and which classmates they have (Coleman, et al., 1966). The impact of which school a student attends differs between countries (Rolfe, Strietholt, & Yang Hansen, 2021), but on average, about 20 percent of the variance in achievement can be attributed to which school a student attends (Opdenakker & Van Damme, 2006). Most of this variance is, in turn, related to class-level differences (Dreeben & Barr, 1988), which in everyday language translates to the different educational opportunities our children get, depending on which school or class they are assigned to.

1.1. The rationale of the thesis

Understanding why these differences occur and why the effects differ between different school systems is of great importance to educational policy. If we want to pursue higher levels of educational equity and reduce achievement gaps between students, then we need knowledge about the mechanisms generating educational inequality. This summons educational research to bridge the gap and enhance our theoretical understanding and
operationalization of peer effects and compositional effects in education. This is the main aim of this thesis.

1.2. Key concepts

This thesis orbits around some key concepts. A few have already been introduced. As mentioned, the term peer effects refers to the ‘influence a student’s classmates has on his educational attainment’ (Arnott & Rowse, 1987). These effects can be both direct or indirect, and endogenous or exogenous. Endogenous peer effects refer to effects where a student’s behaviour (or achievement) is a function of the group’s behaviour, such as disruptive or disorderly behaviour. Many studies on endogenous peer effects therefore concern direct effects.

Exogenous peer effects, on the other hand, are contextual, referring to effects where a student’s behaviour is a function of the group’s characteristics, such as peers’ socio-economic status (SES), gender or immigrant background. Many (but not all) studies on exogenous peer effects therefore concern indirect effects.

Already in this distinction between direct and indirect peer effects, we can catch a glimpse of a probable cause for the ‘remarkably little consensus’ on the nature and size of school compositional and peer effects, as mentioned by Thrupp, Lauder & Robinson (2002). What kind of peer effects could be caused by the mere background of a student’s peers? Are there few or many? How do they function? It is important to bear in mind that peer effects can be both positive and negative. Having disruptive peers most often has negative effects on student achievement, while having high-achieving peers most often is related to positive effects.

While the term ‘peer effects’ is more commonly used by educational economists, researchers in educational sciences more frequently use the terms compositional or contextual effects. These terms include both direct and indirect peer effects. Some researchers use all of these terms (including peer effects) synonymously, while some argue that they should be separated (Wilkinson, 2002). This will be further discussed in this thesis.
As mentioned, the main aim of this thesis is to make a theoretical contribution to research on peer effects in education. The main theoretical framework used is the **Frame Factor Theory (FFT)**, which is an educational theory developed in Sweden during the late 1960s and early 1970s. The theory was developed to enhance our understanding of what actually happens within ‘the black box’ of classroom instruction. The theory questions the idea that classroom teaching is a result of teachers’ conscious determination of the teaching. Instead, the theory predicts that teachers’ instruction can only be done within the limits set by the teaching frames. These frames are the curricula, the time needed for students to learn the curricula, and the time available for instruction. According to the theory, these frames will steer and limit what the teacher can do in classroom instruction in a certain class.

The FFT is an educational theory on classroom instruction and school learning. The aim of providing classroom instruction in schools is to provide students an opportunity to learn the curricula. This perspective – **Opportunity to Learn (OTL)** – has developed over time into a framework for educational measurement within educational research. The framework aims to enhance our understanding and improve our measurement of the characteristics of the instruction that different students meet in school. Such characteristics can be content covered in instruction, the time of content exposure in instruction or the quality of instruction. Given this, the OTL perspective has become central to our understanding of educational attainment and equity.

### 1.3. Research aims

The aim of this thesis is to make a theoretical contribution to the field of peer effects and compositional effects in education. The main argument that will be investigated in the thesis can be formulated as follows:

(i) The Frame Factor Theory (FFT) can be adopted as a theory on peer effects in education.

(ii) The FFT predicts that classroom practice, content coverage and study environment to some extent are dependent on class composition. These factors are mechanisms that generate peer effects and compositional effects on student achievement.
(iii) Since Opportunity to Learn (OTL) investigates differences in educational provision, for instance through content coverage, there is a theoretical overlap between OTL and FFT.

(iv) By premises (i–iii): A theoretical synthesis of the FFT and OTL would hypothetically create a suitable theory with predictive power on peer effects and compositional effects within classroom instruction.

These premises and the complete argument will be tested in this thesis. Since the argument is a deductive argument, the argument can only be true if and only if the premises are true. Therefore, each premise could partly be interpreted as a research question; ‘is it possible to verify each premise?’ The thesis is composed as a compilation thesis, with an integrative essay and three original empirical studies. The different premises are tested in the enclosed articles, and the aim of the integrative essay (the so-called kappa) is to try to establish whether or not the main argument in the thesis is true. Each original study, and the findings, will therefore not be presented in full detail in the integrative essay. Some readers can thereby find it helpful to first read each article, before reading the integrative essay.

Study I examines the possible relationship between disruptive classroom behaviour, school composition and students’ outcomes from an FFT perspective, using Swedish school administrative data (premise (ii) above). The study contrasts previous research, which focuses on the direct peer effect on an individual student’s achievement and not how possible peer effects arise from the interaction between class composition, pupil-pupil interactions and teacher-pupil interactions. In contrast, Study I focus on the interaction between the group (the class) and the teaching process, as predicted by the FFT.

The study verifies a significant relationship between schools’ pupil composition and disruptive behaviour. Disruptive classroom behaviour is also related to the schools’ results measured as the mean GPA. The analysis indicates that some of the original effects of schools’ pupil composition are in fact mediated through disruptive classroom behaviour. Disruptive classroom behaviour is thereby a limiting factor in the teaching process, as predicted by the FFT.
Study II tests a hypothesis that the FFT can be adopted as a theory of peer effects in education (premise (i) above). From the FFT, a multi-level model of classroom instruction is derived, operationalized and tested empirically on Swedish TIMSS data (Trends in International Mathematics and Science Study) of 2015. The analysis reveals that a significant share of the between-class variance in student achievement can be explained by the steering and limiting effect the so-called ‘steering criterion group’ has upon teachers’ instruction. Thus, the study verified the prediction that some peer effects seem to arise from the steering and limiting effect that class composition has upon instruction.

Study III discusses the theoretical overlap of the FFT and OTL and proposes a theoretical synthesis of the two (premises (iii) and (iv) above), where the FFT prediction of class compositional steering and the limiting effects on teachers’ instruction are combined with the OTL perspective of inequalities in educational provision, operationalized as content coverage and instructional quality. From the synthesized model, an operationalized model is derived and tested empirically on Swedish TIMSS data of 2011. The results verify that class composition is related to the presence of limitations on instruction and content coverage, both of which affect opportunities to learn.

These three studies, when taken as a whole, provide indicial evidence for the argument in the thesis. One mechanism generating peer effects in education is the steering and limiting effect of class composition on teachers’ instruction. The student body within the class is both an enabler and a limiter of what can be done in class. The class average level of prior knowledge will to some extent steer and limit what content the teacher can cover in instruction, at what pace and at what depth.

1.4. Disposition of the thesis

The proceeding content of the thesis is organized as follows. In Chapter 2, a short overview will be given of the educational landscape and history of school reform in Sweden. Through this, an understanding of the context in which the FFT was formulated will be enabled. After a presentation of previous research in Chapter 3, the FFT will be presented and discussed in Chapter 4, alongside further adequate theoretical perspectives for the thesis.
The methods and data used in the thesis will be discussed in Chapter 5, followed by a presentation of the findings of the empirical studies and an integrative discussion of conclusions, contributions to the field and recommendations for future research in Chapter 6. Finally, the empirical studies are presented as appendices.
2. Background – The Swedish setting

This thesis is situated in the Swedish school setting. Together with its neighbouring Nordic countries, the Swedish model for comprehensive schooling stands out in an international perspective with low between-school variance in achievement, small gaps in opportunity to learn and high levels of equity (Liu, Van Damme, Gielen, & Van Den Noortgate, 2015; Holmlund, 2020; Rolfe, Strietholt, & Yang Hansen, 2021). Understanding the impact of these conditions is of utmost importance when interpreting the empirical findings within the thesis. Even though the thesis is delimited to peer effects in the Swedish context, both theory and the empirical modelling are connected to the international literature on peer effects.

2.1. Educational reform in Sweden

There is one question within Swedish educational policy that has dominated the political discussion since the end of the 19th century, and that is the question of educational differentiation (Marklund, 1985). The question of educational differentiation in Sweden goes beyond international synonyms such as ‘tracking’ or ‘streaming’. In the Swedish context, these should rather be understood as subsets of the much bigger issue of ‘educational differentiation’, which to some extent also would include a contemporary discussion of segregation following school marketization. In the Swedish context, the question of educational differentiation relates to social democratic ideology and visions for how to reform Sweden mainly during the post-war era (Marklund, 1985).

At the end of the 19th century, Sweden had a parallel school system with ‘folkskola’ [public schools] and ‘läroverk’ [higher secondary schools]. What sort of school a student would attend was mainly determined by social class and parental status. During the post-war era, the social democrats sought to reform the school system. The vision was a comprehensive compulsory school for all children, regardless of social class and background, and that was characterized by high levels of equity, both in input and output (Beach, 2018).
The reform agenda was not without controversy. Both the right-wing political parties and many teachers were sceptical. Were such reforms even possible? How would teachers be able to teach such heterogeneous classes (Marklund, 1985)? To answer some of these questions, large trials were conducted. The most famous of these was the ‘Stockholm trials’ in the 1950s. In these trials, Stockholm was divided into two parts: the municipalities in the north and the municipalities in the south. The students in the north acted as control groups since they were to continue within a differentiated system. The students in the south acted as the experimental group and were organized into comprehensive heterogenous schools for all students.

The most influential analysis of data from the trials was presented in 1962 (Svensson, 1962). In the analysis, Svensson divided the students into three groups: (1) Those who had attended positively segregated schools (higher secondary schools in the north of Stockholm, control group), (2) those who had attended comprehensive schools (comprehensive schools in the south of Stockholm, experimental group) and (3) those who had attended negatively segregated schools (public schools in the north of Stockholm, control group). With controls for student background, aptitude and prior knowledge, Svensson found that positively segregated students in the north (Group 1) achieved the highest results and the negatively segregated students (Group 3) the lowest. Interestingly, he found that, first of all, the differences between the first and second groups were decreasing as the students got older. And secondly, he found that those who gained the most from attending school with more higher achieving peers in comprehensive classes were those who had the weakest aptitude and prior knowledge. Politically, the results were long-desired by the government.

The interpretation of these results was later contested by Urban Dahllöf. In a re-analysis of Svensson’s data, Dahllöf (1967) first found that the tests administered were far too elementary. The tests failed to capture the full width of the knowledge of positively segregated students in the control group. Secondly, Dahllöf found that it was impossible to analyse these results without including process data from the instruction sequences in the schools. When doing so, Dahllöf found that the mathematical content that had taken the students in the experimental group three years to learn had taken the positively segregated students in the control group only one year to learn. In reference to
Carroll’s ‘Model of School Learning’ (1963), Dahllöf interpreted these differences as a result of students’ prior knowledge. Students’ prior knowledge had, in this case, enabled teachers to cover the content faster. Based on these insights, Dahllöf initiated the formulation of a pedagogical theory – the Frame Factor Theory. The theory will be further discussed in Chapter 4.

The process with the trials ended in several different political compromises, and in 1962, the new comprehensive compulsory primary and lower secondary school for all was created. Since then, further reform has been undertaken. Successively, different types of differentiation that was a result of the compromises – tracking, streaming and grouping – were abolished (Marklund, 1985). The last formal rules for tracking students were abolished in the curricular reform of 1994 (Sund, 2013).

A giant shift in Swedish school policy occurred during the 1990s when the right-wing government introduced privatization and marketization in the school system. These reforms had been preceded by decentralization and deregulation of the school system during the 1980s and 1990s. This is of utmost importance for understanding the effects that privatization would have in Sweden. It could be argued that the previous deregulation of the school system – with abolished rules for teacher qualifications, class sizes and less central control of other instructional conditions – created a specific context within which the privatization and marketization were implemented and developed (Bäckström & Isaksson, 2022). Taken together, the reforms of the late 1980s and early 1990s shifted the focus of the Swedish school system from more central planning and regulation of input factors to a more New Public Management-oriented system focusing on freedom of choice, market-oriented competition and measurement of output (Englund, 1996; Larsson, 2011).

These reforms have had a major impact on the Swedish school system. It has marked a shift in teachers’ relation to their mission, their students and pathos (Fredriksson A., 2011), it led to increased variance in teachers’ working conditions and psychological strain (Kjellström, Almquist, & Modin, 2016) and it has re-introduced a sorting and segregation of students reminiscent of those at the end of the 19th century (Holmlund, et al., 2014).
It is important to emphasize that the increased school segregation in Sweden is mainly driven by increased residential segregation, which is a result of the increased variance in income distribution and other inequalities in society. The introduction of school choice and private schools account for about 20 percent of the increased school segregation from the early 1990s (Holmlund, et al., 2014). The increased school segregation in Sweden has widened the achievement gap between schools (Holmlund, et al., 2014; Skolverket, 2018).

2.2. Achievement, equity and peer effects in education

From an international perspective, the post-war educational policy in Sweden created a school system that had a heterogenous school composition regarding students’ backgrounds, small between-school OTL gaps and highly equitable results (Schmidt, Burroughs, Zoido, & Houang, 2015; Rolfe, Strietholt, & Yang Hansen, 2021). Since the 1990s, this has changed, and no other OECD country has decreased both the levels of achievement and equity in the same manner as Sweden (Skolverket, 2014). To date, Sweden is closing in on the OECD average in several of these indicators (Skolverket, 2019; 2020). This development for Swedish elementary schools is illustrated in Figure 1.

A driver of increased between-school variance in Sweden is the increased segregation (Holmlund, et al., 2014; Yang Hansen & Gustafsson, 2019) and the increased differences in students’ prior knowledge due to the vast immigration to Sweden during the late 20th century (Skolverket, 2018; Gustafsson & Yang Hansen, 2018; Katz & Österberg, 2020). These changes have spurred a debate on the importance of peer effects in Sweden, a debate that is both academic (Gilljam & Persson, 2010; Gustafsson J.-E. , 2010) and political (Heller-Sahlgren, 2015; Boguslaw, 2018). This debate must therefore also be seen as the context in which this thesis is written. Given this, the research is delimited to peer effects in the Swedish context, but both theory and the empirical modelling are connected to the international literature on peer effects.
Figure 1: Share of explained total variance in students’ expected GPA in year 9 (lower secondary school)

Remark: The figure was originally reported in Holmlund et al, 2014, Figure 5.18, p. 96. Adjusted school effect is adjusted for residential segregation.
3. Previous research

For more than 50 years, there has been extensive research conducted into peer effects in education, with many researchers pointing to the Coleman Report (1966) as a starting point in the US (Barr & Dreeben, 1977; Zimmerman, 2003). The combined literature on peer effects is vast, but researchers have drawn attention to the fact that ‘there has been much less investigation of the ways in which they [peer effects] operate in schools or of their effects on school functioning or pupil progress’. (Rutter & Maughan, 2002, p. 469). Also, several researchers have taken notice of the lack of consensus within the field. Thrupp, Lauder & Robinson (2002) note that ‘there is remarkably little consensus over the nature and size of school compositional and peer effects’ (abstract). One reason for this lack of consensus is that school compositional effects have been ‘inadequately theorised and poorly operationalised’ (p. 484).

As noted in the last paragraph, there is a difference in terminology. In educational research, the topic of interest has been labelled ‘compositional’, ‘contextual’ or ‘structural’ effects (Dreeben & Barr, 1988), while foremost economists speak of ‘peer effects’ (Arnott & Rowse, 1987; Epple & Romano, 2011; Yeung & Nguyen-Hoang, 2016). The issue at the heart of the research remains the same: How important are classmates for individual student’s achievement? However, it is possible that the differences in terminology also bear witness to significant differences in the understanding of the effects in focus, which possibly could contribute to the ‘remarkably little consensus’ within the field.

3.1. Compositional effects and peer effects in education

The terms ‘peer effects’ and ‘compositional effects’ often tend to be used synonymously, but Wilkinson (2002) has suggested a distinction between the two where ‘peer effects’ should relate to the more direct effects that peers have on each other (such as normative, competitive and disruptive factors) and ‘compositional effects’ should relate to the more indirect effects of peers’ aggregated characteristics.
Especially researchers in the economics of education have upheld distinctions of this sort. In this research, it is common to distinguish between endogenous and exogenous independent factors generating peer effects. The former, endogenous factors, concerns individuals’ behaviour and assumes that individuals’ behaviour will be related to the behaviour of the group. They are therefore often measured as what people do, how they behave and what they are capable of. The latter, exogenous factors, are measures of their characteristics, and it is assumed that an individual’s behaviour will be associated with the characteristics of the individual’s group (Epple & Romano, 2011). Studies of endogenous factors can thereby be studies utilizing survey data, observational data or register data on aptitude or achievement, while studies of exogenous factors most often utilize data on students’ educational, family and immigration background.

As mentioned, many researchers point to the Coleman report (1966) as a starting point for this strand of research (Barr & Dreeben, 1977; Zimmerman, 2003). Since then, a myriad of studies have reported similar results on these peer group effects (Hoxby C., 2000; Hoxby & Weingarth, 2006; Ewijk & Sleegers, 2010; Lin X., 2010; Hanushek & Woessmann, 2017). For instance, Szulkin & Jonsson (2007) reported that if a lower-secondary school class in Sweden had more than 40 percent foreign-born students, the results for the same group decreased by an average of 0.20 standard deviations. Similarly, Brunello & Rocco (2013) studied 19 countries participating in PISA and found that if the share of immigrants in secondary schools were to double, the native students’ achievement would decrease by 1.0–3.4 percent. Similar results can be found in Betts (1998), Gould, Lavy, & Paserman, (2009), Jensen & Würtz Rasmussen (2011) and Chetty, Hendren, & Katz (2016).

There is also a vast number of studies using endogenous factors to study peer effects. In a meta-analysis of 53 studies between 1980–2014, Yeung & Nguyen-Hoang (2016) found a small but significant effect. They concluded that ‘there is a genuine empirical endogenous effect of peers on student outcomes. Better peers result in better outcomes’ (p.45). Similar findings were reported by Gutiérrez (2023). Utilizing data from Chile, Gutiérrez analysed the extent to which socioeconomic and academic classmates’ characteristics were associated with student attainment. He found a small stable effect of increases in peers’ academic performance but almost no effects associated
with changes in the classmates’ socioeconomic status. Gutiérrez’s findings thereby indicate that the main mechanisms of peer effects are concerned with peers’ prior knowledge, not their background.

Several other studies support this interpretation. For instance, Burke & Sass (2013) report a non-linear effect of having high- or low-ability peers. Meschi & Pavese (2023) report that immigrant students are detrimentally affected by the share of very low achievers in the classroom, in comparison to native students who mostly are influenced by the average quality of their peers. Similar findings can be found in Beckerman & Good (1981), Hanushek, Kain, Markman, & Rivkin (2003), Schindler Rangvid (2007), Imberman, Kugler, & Sacerdote (2012) and Chen et al. (2020).

There are also several studies on the effects of peers’ behaviour (Bru, 2006; Smith & Reimer, 2018; Duesund & Oedegaard, 2018; Agasisti, Avvisati, Borgonovi, & Longobardi, 2018). Lavy & Schlosser (2011) studied the effect of disruptive classroom behaviour in Israeli schools. They concluded that a larger share of girls in the class was related to positive peer effects for all students. The effect was mediated through less violence, improved relations and less teachers’ fatigue. In another study, Lavy, Paserman & Schlosser (2011) reported evidence that an increase in pupils who have been retained for one year was associated with a decrease in the new classmates’ achievement levels. The effect was mediated by detrimental effects on teachers’ pedagogical practices, increased violence and classroom disruptions. Similar results can be found in Weinert, Schrader, & Helmke (1989), Lazear (2001), Aizer (2009), Bertrand & Pan (2013), Cheema & Kitsantas (2014) and Grøne Kristoffersen et al. (2015).

Returning to the distinction between ‘peer effects’ and ‘compositional effects’ made by Wilkinson (2002), there are reasons to question whether or not this is firstly a feasible and secondly a coveted distinction. Is it possible to withhold the distinction and separate the effects from one another (Dumay & Dupriez, 2007)? And if so, is it theoretically desirable? Several studies have indicated that there is a relationship between exogenous and endogenous factors, where the latter seems to depend on the former (Lavy & Schlosser, 2011). The more diverse the class is, the larger the share of low-performing students there is and the larger the negative effect there is on achievement
(Grøne Kristoffersen et al., 2015). The meta-understanding of this phenomenon must be that there is also a relation between compositional effects and peer effects. The probability of ‘direct’ peer effects, in Wilkinson’s terminology, seems to be much greater in classrooms with more ‘indirect’ compositional effects (Lavy, Paserman, & Schlosser, 2011).

In reference to this, some researchers have claimed that a model of compositional and/or peer effects must capture the more complex relations within classrooms (Dumay & Dupriez, 2007; Liu, Van Damme, Gielen, & Van Den Noortgate, 2015). One such example is found in Bourke’s (1986) study of the relations between class size, teacher practices and student achievement, but a more comprehensive model for this was developed by Thrupp (1999) in his ‘School Compositional Hypothesis’ (SCH).

3.2. Models of compositional effects

3.2.1. The school compositional hypothesis

In 1999, Thrupp contested the up until then very influential ‘school effectiveness’ research. One of the main arguments of Thrupp was that the school effectiveness research had a socially decontextualized view on school quality (Dumay & Dupriez, 2007). In this strand of research, compositional effects in education – such as those found by Coleman et al. (1966) – tended to be viewed as an external problem which must be controlled for in the search of the more important determinants of student achievement (Dumay & Dupriez, 2007). Thrupp instead argued that school intake, school mix and compositional effects must be seen as parts of the whole school’s functioning. Thrupp summarized this in the ‘School Compositional Hypothesis’ model (SCH). The essence of this model is built upon the seminal work of Barr & Dreeben (1983) and Dreeben & Barr (1988), and the notion that compositional effects mainly operate through a nested series of hierarchical layers, where school, class and group compositional effects are intertwined, all affecting each other, and in the end also affecting the students’ achievement. Or put another way: The school compositional effect is not only a direct effect on student achievement but also an indirect effect mediated by managerial, pedagogical and psychosocial processes.
In his study, Thrupp (1999) found evidence that high SES schools had a more academic school program, more able and motivated students, more demanding classes and more qualified teachers than low SES schools. At the psychosocial level, he found higher levels of prior attainment, higher goals, less violence and less disrupting behaviour in the high SES schools. Within classes in these schools, he also found that teachers had to focus less on discipline management, there were higher levels of student compliance and the teachers could use a wider array of teaching practices. Similar results can be found in Weinert, Schrader, & Helmke (1989), Garrison (2004) and Blatchford et al. (2007).

3.2.2. Conceptual model of peer influences on learning

In the same themed issue on peer effects in education, where Wilkinson proposed a distinction between peer effects and compositional effects (2002), the authors synthesize the findings in their review on peer effects and compositional effects into a ‘Conceptual model of peer influences on learning’. The model is depicted in Figure 2.

As the figure indicates, the direct peer effects of both inter- and intrapersonal character are intertwined in the model, affected by processes on other levels and within other hierarchical layers within schools. In their explanation of the model, the authors emphasize that the three planes at the school, class and group levels represent the manners of influences on learning associated with each level. The vertical arrows in the figure represent compositional effects, including peer effects. The authors also hypothesize that the majority of these effects are indirect.
In a study on PISA data from 28 countries in 2003, Liu et al. (2015) took parts of the model to empirical test. Since previous studies had found an overlap between school compositional effects and the effects of school processes, the aim of the study was to further examine if school processes in fact mediated the relationship between school composition and school achievement. They found that it did. They found varying school compositional effects across countries, which confirmed previous findings. They found three meaningful mediators from school climate factors. These were the school disciplinary climate, students’ positive behaviour and student morale. Each of these significantly accounted for school compositional effects.

Besides the evidence from Liu et al. (2015), other research has also provided evidence for SCH-related models. Such results can be found in Opdenakker & Van Damme (2001; 2006) and Dumay & Dupriez (2007).
3.2.3. The Opdenakker model of compositional effects

The Opdenakker model of school compositional effects is very similar to the SCH model (Opdenakker & Van Damme, 2006). In the model, it is assumed that a student’s effort and achievement are related to characteristics of the student himself as well as to characteristics of classes and schools. Effort and achievement are assumed to be influenced in two ways: by their influence on class characteristics and by their influence on the relationship between class characteristics and achievement. In turn, school characteristics are divided into three main types:

- characteristics referring to the composition of schools (student population, teaching team and school leader),

- characteristics referring to the school practice (educational framework, organization and management, work and learning climate) and

- context characteristics (e.g. denomination, school size, study programme offerings). (Opdenakker & Van Damme, 2006, p. 88f)

The full Opdenakker model is depicted in Figure 3. When testing the model empirically, Opdenakker & Van Damme (2006) found that of the variance in mathematics achievement that could be attributed to the school level (ICC 19 percent), group composition, social and learning climate and opportunity to learn could explain 90 percent.
3.2.4. **TIMSS model of potential educational experiences**

In their seminal work, Barr & Dreeben (1983) argue that how schools work must be understood in the same sociological manner as we understand other organizations: decisions, rules and given conditions at different levels affect other levels within the organization. As discussed, this perspective has been of utmost importance for subsequent research on school compositional effects (Thrupp, 1999; Wilkinson et al, 2002).
In the same manner, this perspective is of great importance for IEA (International Association for the Evaluation of Educational Achievement) in general, but for TIMSS (Trends in International Mathematics and Science Study) in particular. TIMSS considers curriculum to be a complex and multifaced construct; ‘a delicate web of multi-level relations between aims, actors, and choices across systems, classrooms, and individuals’ (Rolfe V., 2021, p. 45). TIMSS’ understanding of curriculum – and school functioning – is formulated in the ‘TIMSS Model of Potential Educational Experiences’ (Schmidt et al, 1997). It represents a comprehensive model of educational delivery of school systems. Besides a more holistic system perspective than that of Barr & Dreeben (1983), the model integrates important perspectives from Opportunity to Learn (see Chapter 4), such as the differences between ‘intended curricula’ and ‘enacted curricula’. The model also includes compositional effects in instruction. The model is depicted in Figure 4.

![TIMSS Model of Potential Educational Experiences](source: Rolfe, 2021, p. 46 after Schmidt et al., 1997.)
3.3. Towards a theory of peer effects

As Liu et al. (2015) noted, there are to date few studies that investigate how school composition affects student achievement. They write:

To date, many researchers perceive several sources of substantive school composition effects: peer effect, teacher/teaching effect, and facility effect, while others consider them as a statistical artefact, due to model under-specification and/or poorly measured variables […]. Nevertheless, there are no comprehensive theories explaining the mechanism of school composition effects and most researchers treat it as a ‘black box’. (Liu et al., 2015, p. 424; my emphasis)

In reference to the research aim of this thesis (Chapter 1), this is a focal point. The argument for applying the FFT to the issue of peer effects is that it provides a theoretical understanding of some of the mechanisms generating peer effects within classroom instruction and how this is related to class composition and OTL.

It could be argued that the type of synthesized model of FFT and OTL that I will present in Chapter 4 already exists through the different models of compositional effects discussed in Section 3.2. My argument, in line with Liu et al. (2015), is that these models lack appropriate precision and detail in predicting the mechanisms that would generate peer effects and/or compositional effects within classroom instruction (Thrupp, Lauder, & Robinson, 2002; Rutter & Maughan, 2002). With this said, there is no reason to interpret the synthesized theoretical model that I will present in this thesis as contradicting some of the other models discussed. Rather, it should be viewed as a classroom model, subsidiary to (or possibly integrated into) other system-related general models.
4. Theoretical framework

The theoretical framework of this thesis is grounded in the Frame Factor Theory (FFT). Throughout the work with the thesis, both my understanding, implementation and operationalization of the FFT have evolved. In studies I and II, a somewhat ‘original’ version (Dahllöf, 1967; Lundgren, 1972) of the theory was applied. The empirical findings in Study II spurred a discussion on construct validity regarding some predictions of the FFT. This discussion led the forthcoming inquiries to the OTL perspectives in general, and its view on content coverage in particular.

Therefore, in this chapter, I will first discuss each of these theoretical perspectives. Then I will present the synthesized version of the theory that was tentatively discussed in Study II but finalized and applied in Study III.

4.1. Frame Factor Theory

As mentioned in Chapter 2, the early ‘frame factor thinking’ (Dahllöf, 1999) was spurred on during the trials with a comprehensive elementary school reform in Sweden. Dahllöf’s (1967) re-analysis of data from the trials showed that variation in classes’ average prior knowledge and aptitude both steered and limited the instruction in class. In the positively segregated classes (the control group), the teachers could cover the elementary content at a faster pace, generating more time to focus more on advanced content. The mathematical content that had taken the positively segregated students in the north one year to cover had taken the students in the comprehensive schools in the south three years to cover.

At the heart of the FFT is the notion of the relation between the time needed (T_N) for students to learn a specific unit in the curriculum (C_U) and the actual time at hand (T_A), introduced by Carroll in his ‘Model of School Learning’ (1963). According to this model, if the time available (T_A) is less than the time needed (T_N), then students will not learn:

\[(T_N) > (T_A) \rightarrow \neg(CU)\]
Other contemporary models of school learning at that time, such as Bloom’s Mastery Learning (ML) (Bloom, 1974), also shared a common ground with Carroll’s model of school learning. A crucial difference between the FFT, ML and other more psychological models of the educational productivity process (Haertel, Walberg, & Weinstein, 1983) was that the focal point was not the individual student; it took the teachers’ perspective (Barr & Dreeben, 1977; Broady, 1999). This is illustrated through the FFT’s notion of the steering and limiting effect that class composition will have on teachers’ instruction. In an analysis of instruction, the teachers’ practice to some extent must be viewed as dependent on class composition. The teacher is therefore both a dependent and an independent variable in the analysis. Lundgren comments on the matter:

Why have different theories on teaching had so little influence on the teaching process? The answer seems to lie in the fact that these theories are not based on empirical analysis of teaching. Instead, they have built on different ideas of what teaching is. One such idea is that the teacher consciously determines the teaching. We hold that his steering can be done only within the limits set by the teaching frames. (Lundgren, 1972, p. 38f)

Interpreted differently, Lundgren hereby also makes a claim on teachers’ managing of the class: classroom management and classroom instruction is a collective practice. The group will always set a limit for how much instruction can be individualized. Such a note is also given by Barr & Dreeben:

The difficulty with the mastery learning formulation is that it is conceptualized to deal with individuals as if they were receiving tutorial instruction […], when in fact they are members of a classroom receiving group instruction. (Barr & Dreeben, 1977, p. 117f)

They continue:

Bloom treats the management of learners (in particular, the securing of their attention) as distraction pertaining more to administrative expectations and the need for control than to instruction and learning. But it is not distraction; it is an integral part of classroom teaching precisely because classes are social aggregates whose members have diverse individual characteristics and whose conduct must be managed en masse, in groups and individually. Classroom instruction cannot be understood in terms of a model based on congeries of
dyads, and while the management of learners does not usually appear as a problem in tutorial instruction, it does in classroom instruction. (p. 121)

Similar conclusions have been made thereafter. Research by Weinert, Schrader & Helmke (1989) found that the contextual setting the class composition makes up must not be neglected when assessing instructional quality and its effects on student achievement. They found considerable correlations between aspects of instructional practices and classroom composition and characteristics. For instance, where the students in a class had better than average prior knowledge and aptitude, there was less disruptive behaviour, more effective classroom management and more complex questions could be discussed in class. They write:

These findings contradict the widely accepted notion that ‘good instruction’ is exclusively a result of teachers’ aptitudes and skills. Rather, the effects of teacher behaviours tend to be more or less confounded with context factors. Moreover, some context factors (particularly [students’] level of prior knowledge and ability) are not only determinants of teacher behaviour but also regarded as aptitudes determining subsequent achievement outcomes. Student characteristics, such as prior knowledge and ability, obviously can have a direct as well as an indirect impact (via instructional processes) on student achievement. (p. 907)

To date, this FFT perspective can be found in contemporary research, even though it is described in different terms. One example is found in Nuthall:

In order to take responsibility for a class of 25 to 35 students who have different knowledge, skills, interests, and motivations, teachers have to develop ways of managing the class as a whole. It is impossible to focus on the individual learning of any one student for more than very brief periods. (Nuthall, 2005, p. 903)

As mentioned, the FFT claims that the three main frame factors – time needed, time at hand, and the curriculum – will steer and limit the teaching process. The steering and limiting effect arises if the time needed by the class is less than the time at hand because then the teacher must act in some way. The teacher’s possibilities include holding up the entire class, moving on, lowering demands or altering the content covered (Lundgren, 1972; 1973).
The FFT hypothesizes that the presence and magnitude of the steering and limiting effect can be predicted by the so-called ‘steering criterion group’ (SCG). The SCG is defined as the average aptitude of the students between the 10th and 25th percentile of aptitude distribution in class, and the FFT predicts that the teacher will make decisions within the instruction in reference to this group (Dahllöf, 1967). From the teachers’ perspective, the SCG therefore is not a formal group, but despite this, Lundgren’s (1972) empirical micro-studies of instruction revealed an extensive agreement between his statistical definition of the SCG within classes and the teachers’ notion of which students he or she based collective decisions upon regarding instructional progress. Regardless of this, the focal point is not the exact definition of the percentiles (Arlin, 1984) but more the principle that classroom instruction is, and always will be, a collective practice (Barr & Dreeben, 1977; Thrupp, 1999; Nuthall, 2005), and within this practice, the teacher will have to make decisions about how to progress. These decisions will have different effects for different individuals in the class.

According to Lundgren (1972), the most important frame factors are class composition and time available. He writes: ‘The relationship between frames and processes means not only steered by the frames but also limited by them’ (1972, p.13). This means that the frames not only make instruction possible, but they also limit the same instruction.

A contemporary critique of the FFT was that it was far too deterministic (Broady, 1999). This is something that Lundgren commented on already in his early works. The frames will steer and limit the teaching process but only ‘to a certain extent determine the teaching process’ (1972, p.27; my emphasis). Teachers will always have different personalities as a teacher, different didactical approaches and prioritize content in different ways, which will affect both instruction and student achievement. When taking the frame factors as independent variables in an analysis, we could therefore expect an explained variance of between 10 and 25 percent. Thereby, according to Lundgren (1972, p.40), correlations of $r = .30-.50$ could be interpreted as support for the theory.
In Figure 5, an original paradigm of the FFT is depicted. The paradigm comes from Lundgren’s thesis in which he finalized the early version of the theory.\textsuperscript{1} The paradigm shows the relations between the frame factors and the teaching process. As shown, the teaching process will be affected in different ways (right-hand side of the model). First, if the time at hand is less than the time needed, then the teacher most likely will have to spend more time on elementary parts of the curricula, forsaking more advanced content. Also, the greater the variance in class, the greater the risk of boredom for high-achieving students. Second, if the time at hand is less than the time needed, then the theory predicts that teaching practices will be affected. One effect can be how speaking time is distributed and used and what type of dialogue occurs.

\textit{Figure 5: Paradigm of the Frame Factor model}

Source: Lundgren, 1972, p.43

\textsuperscript{1} Urban Dahllöf was Ulf P. Lundgren’s supervisor.
For instance, the theory predicts that the lower level of the SCG in class, the more the students will be left alone with individual work and not taught in whole class instruction (compare with Garrison, 2004; Blatchford, Edmonds, & Martin, 2003; Blatchford et al., 2007).

In contrast to Dahllöf’s re-analysis of data from the Stockholm trials, Lundgren’s thesis (1972) was conducted within a self-designed research project led by Dahllöf. The thesis was based on two empirical studies, one quantitative study in which 47 8th grade classes were followed in a longitudinal setting and one qualitative study where one class was followed during one whole semester. Given the predictions mentioned, Lundgren reports 85 correlations in his thesis. In Figure 6, six of these 85 correlations are depicted as an example. As expected, the figure shows that in classes with a higher level of the SCG, there were more teacher-led discussions in the whole class (r = 0.82) and higher levels of teacher ‘acceptance’ of ideas (0.67).

![Correlation Diagram](image)

**Figure 6: Correlations between SCG and classroom communication**

Remark: Excerpt of Lundgren, 1972, fig.17:1-2, p.302f. Dotted lines refer to communication in ‘cognitive dimensions’ and solid lines to ‘affective dimensions’.
At the same time, teachers in these classes had to spend less time managing disruptive behaviour (-0.73), and there were fewer monologues where the teacher had to inform the class on different matters of instruction (-0.76).

These empirical findings were well in line with the predictions of the FFT, and Lundgren interpreted them as evidence of the theory. He concluded:

> We conclude that the composition of a class clearly influences the shape the teaching process takes, and, as a matter of logic, also its outcome, and that this result is dependent on the fact that the teacher using traditional classroom instruction adapts to a considerable extent the teaching according to the degree of ability in a certain group of pupils. These pupils set the pace of the teaching and decide the amount of information the teacher can give. (Lundgren, 1972, p. 338f)

Many of the early FFT studies were conducted within mathematics due to its inherent cumulative nature. Thus, further studies were conducted on other subjects (Swedish, for instance), and similar results were found (Dahllöf, 1971; 1999).

### 4.2. Opportunity to learn

In their seminal work, Coleman et al. (1966) investigated the equality of educational opportunity within the US. Alongside evidence of poorer teacher quality in minority schools, one main finding was that pupils’ achievement was related to the educational backgrounds and aspirations of the other students in school. Contemporary US studies have found similar results. In a study of US TIMSS data from 2011, Morton & Riegle-Crumb (2020) found that teachers in predominantly black schools reported significantly lower levels of time spent on advanced mathematical content than teachers in predominantly non-minority schools did. This indicates that students in these different schools get different opportunities to learn, which to some extent could explain the achievement gap between students (Schmidt, Siwen, & Houang, 2021).

Given this, Opportunity to Learn (OTL) has become an important perspective when studying educational inequality (Tate, 1995; Schmidt, Burroughs, Zoido, & Houang, 2015). As the name implies, OTL concerns students’
opportunities to learn the intended curricula. Early OTL research therefore focused on studying the alignment between the intended curriculum and the enacted curriculum\(^2\) by teachers in classrooms (Elliot & Bartlett, 2016).

This means that OTL also shares the common ground in Carroll’s (1963) model of school learning with the FFT and the ML (Kurz, 2011; Elliot & Bartlett, 2016), or as Carroll (1963) puts it: ‘the learner will succeed in learning a given task to the extent that [the learner] spends the amount of time that [the learner] needs to learn the task’ (p. 725, compare Section 4.1). Therefore, in Carroll’s model, student learning is a function of time spent (compare Wayne & Walberg, 1980) and, given this, schools have to provide students with the amount of time needed for the specific learning objective. Otherwise, they will not have been given an opportunity to learn the intended curricula.

The unit of analysis for most OTL research is therefore the classroom, and the main interest is focused on studying those ‘inputs and processes within a school context necessary for producing student achievement of intended outcomes’ (Elliot & Bartlett, 2016, p. 1). The research includes questions of whether students are given the same access to and possibilities for learning. This is often operationalized as instructional time and content (Wayne & Walberg, 1980; Barr, 1988; Porter, 2002), but the variety of OTL measurements is great (Rolfe V., 2021).

Going beyond this, both OTL, the FFT and the ML share the perspective that the time allocated for instruction also de facto must be used for the given purpose; actual time spent on instruction almost always subside the time

\(^2\) Compare with ‘TIMSS Model of potential educational experiences’, Figure 4, p. 20.
allocated for instruction (Kurz, Elliott, Kettler, & Yel, 2014; Elliot & Bartlett, 2016). This perspective has been labelled as ‘time-on-task’ (Anderson, 1981; Karweit, 1984) or ‘engaged time’ and is central – pronounced or non-pronounced – for the named theoretical perspectives as well as for research into endogenous peer effects such as disruptive behaviour (Lazear, 2001; Lavy & Schlosser, 2011).

One of the first comprehensive OTL frameworks was presented by Stevens (1993). Stevens’ framework rested on studies identifying four groups of variables that seemed to have a powerful influence on teacher instructional practices and student learning. These were:

1. Content coverage (measuring whether students cover the core curriculum),
2. Content exposure (measuring allocated time and time-on-task),
3. Content emphasis (measuring for which students different content is emphasized and in what manner (lower/higher order skills)) and
4. Quality of instructional delivery (measuring teaching practices).

The framework of Stevens rests upon the crucial concepts of time for instruction (Carroll, 1963), content coverage (Husén, 1967) and the quality of instruction (Walberg, Pascarella, Haertel, Junker, & Boulanger, 1982; Weinert, Schrader, & Helmke, 1989), all inalienable perspectives when studying student achievement. According to Elliot & Bartlett (2016), there is a lack of direct empirical evidence for Stevens’ framework, but several contemporary researchers still rest upon her framework (Kurz, Elliott, Kettler, & Yel, 2014). In a study of the relations between OTL and science achievement, Wang (1998) utilized Stevens’ framework and found that content exposure was the most significant predictor of student achievement in written tests and quality of instructional delivery in hands-on test scores.

In the US, OTL research is prominent, with perspectives dating back to the Coleman report and with Stevens’ (1993) framework still at play (Kurz A., 2011; Kurz, Elliott, Kettler, & Yel, 2014; Elliot & Bartlett, 2016; Wang, Liu, & Leung, 2022). Contemporary studies focus on the same issues as Coleman did, for instance, regional differences in the provision of education (Schmidt,
Cogan, Houang, & McKnight, 2011; Reeves, 2012; Saw & Agger, 2021) but also differences in educational provision and opportunities to learn between minority and non-minority students.

In a recent study of the latter, Morton & Riegle-Crumb (2020) utilize US TIMSS 2011 data to study the relation between school ethnic composition and content coverage in algebra. They found that the teachers’ reported time spent on algebra and advanced 8th grade content was lower in schools that were predominantly black than in those that were not. Under control for school characteristics (including proxies for SES), teacher qualifications and students’ academic performance, they found that in schools that were predominantly black, teachers on average spent 72 percent of class time over the year covering algebra and advanced content, while the corresponding share in predominantly non-minority schools was 82 percent.

From a perspective of achievement and equity, OTL gaps are problematic since there is a consistent body of OTL literature indicating a positive relation between OTL and achievement. In a summary of the meta-analysis on school-effectiveness variables, Scheerens (2017) reports that OTL is a consistent predictor of achievement with average standardized effect sizes (Cohen’s d) of 0.18–0.88.

Further complexity is added to this issue since several studies have found relations between OTL gaps and school composition (Rolfe V., 2021), such as in the case of Morton & Riegle-Crumb (2020). In the international literature, studies have shown that OTL alone explains less than 10 percent of between-school differences in achievement on average across OECD countries, but in conjunction with SES, it explains nearly 30 percent (Lafontaine, Baye, Vieluf, & Monseur, 2015; Rolfe V., 2021). According to Guiton & Oakes (1995), these effects could be a result of teachers [expecting] mixed and predominantly minority classes to have less previous exposure to these topics, warranting treating more topics as new material, yet simultaneously holding lower expectations for students mastering the content. (p. 330)

Another recent example of the same issue from Sweden is a study by Klapp & Jönsson (2020). The study is based on a survey of Swedish 9th grade
students in lower secondary schools (on average 15 years old), which was intended to detect shortcomings in schools’ provision of support to students at risk of not reaching a passing grade in one or several subjects. They report evidence that low-achieving students perceive that, to a large extent, they are provided with ‘simplifying support’, meaning that the teachers lower the standards and their expectations of them, thereby limiting these students’ opportunities to learn.

It is significant for both the Klapp & Jönsson (2020) and Morton & Riegle-Crumb (2020) studies – and also Guiton & Oakes (1995) – that in reference to the theoretical perspectives of the FFT, they never discuss why these results were found. The mechanisms generating the phenomena are not discussed. Do teachers lower their expectations due to malice, or are there any other mechanisms at play?

From an FFT perspective, this phenomenon must be understood as part of the compositional steering and limiting effects on teachers’ instruction. Morton & Riegle-Crumb’s (2020) study facilitates a possible interpretation: If the curriculum states that a class is supposed to be taught advanced algebra, what should the teacher then do if a very large share of the students in class do not master multiplication and division? Since multiplication and division are prerequisite knowledge for algebra and calculus, the teacher will have to start by teaching multiplication and division. Is this a result of teacher malice or actual circumstances?

3 According to the Swedish school law, any student at risk of not reaching the passing grade should be given extra support by the school.
I argue that this discussion identifies a need for a theoretical synthesis of the FFT and OTL.

4.3. The need for theoretical synthesis

It is important to note that the results reported by Morton & Riegle-Crumb (2020) are not expected or predicted by the FFT. The theory does not predict that teachers will make a difference between students by their race or ethnicity, only that their practices to some extent will be steered and limited by the class level of prior knowledge and average aptitude. Having said this, if minority status is in fact a proxy for lower prior knowledge in predominantly black schools in the US, it could be argued that this effect is expected by the FFT. To some extent, this interpretation seems plausible since there is strong evidence of the relationship between students’ background and their academic achievements (Sirin, 2005).

No matter what, the results from Morton & Riegle-Crumb (2020) highlight the need for a theoretical (and methodological) synthesis of the OTL perspective and the FFT when studying educational opportunity and peer effects in education. The FFT predicts that there will be a difference in coverage of elementary and advanced content between high and low-performing classrooms. Teachers in low-performing classrooms will have to spend more time on elementary parts of the curricula in order to get the majority of the class to master the core elements (Lundgren, 1973). This means that the OTL operationalization of Morton & Riegle-Crumb (2020) – but in essence the OTL framework of Stevens (1993) – could be incorporated into an FFT model predicting peer effects in education, effects mediated by the compositional effect on teachers’ instruction. While enhancing the precision of the FFT, the synthesis would at the same time offer a theoretical explanation to some of the empirical results reported in much of the OTL literature, for instance, those of Morton & Riegle-Crumb (2020) and Klapp & Jönsson (2020).

As mentioned in Section 3.3, it could be argued that synthesized models of such already exist. Thrupp’s SCH model was discussed (Thrupp, 1999; Dumay & Dupriez, 2007) alongside the Opdenakker model, educational productivity models such as that of Walberg et al. (1982) and the
comprehensive ‘TIMSS Model of Potential Educational Experiences’ (Schmidt, Raizen, Britton, Bianchi, & Wolfe, 1997; Rolfe, Strietholt, & Yang Hansen, 2021). My argument, in line with Liu et al. (2015), is that these models lack appropriate precision and detail in predicting the mechanisms that would generate peer effects and/or compositional effects within classroom instruction (Thrupp, Lauder, & Robinson, 2002; Rutter & Maughan, 2002).

The argument for this synthesis of OTL and FFT is mainly to merge the perspective of content delivery within OTL with the class compositional steering and limiting perspective of the FFT, into one common theoretical perspective on peer effects in classroom instruction, which can be modelled and empirically estimated. The estimation done within this study should thus be viewed as a classroom model, by necessity not contradicting some of the other models discussed. Rather, it should be viewed as a classroom model, subsidiary to – or even possible to integrate into – other system-related general models.

In Figure 7, the synthesized theoretical model applied in Study III is presented. The model illustrates the predicted relation between the frame factors (Lundgren, 1972; 1973) and student achievement as well as how this is mediated through an indirect effect via teaching practices (Stevens, 1993). The expected relations between class compositional variables (such as students’ average socio-economic background (SES)) and the frame factor of students’ ‘time needed’ on the one hand and ‘teaching practices’ on the other, are marked by one arrow each. Controlling the effects in the model for class compositional characteristics is marked by the dotted arrow to students’ achievement. In contrast to this theoretical model, the final operationalized model is reported in Study III, found in the appendices.
Figure 7: The synthesized theoretical model of peer effects on the between-class level
5. Methods and data

The international literature on peer effects is dominated by research applying different quantitative methods such as variants of multilevel multivariate regression analysis. This is most likely due to the nature of research aims and questions within the field. The research is often aimed at establishing relations between – or even causal effects of – class composition, different classroom factors and student achievement, why quantitative research methods are suitable.

This thesis is no exception. As discussed in Chapter 1, the aim of the thesis is to make a theoretical contribution to the field of peer effects and compositional effects in education. The aim therefore must be understood in relation to previous research into peer effects and opportunity to learn, the theoretical perspectives in this research, its models, and empirical findings. Also, since the aim of the included studies in the thesis is to explore different predicted relations between school- and classroom factors and student achievement, according to the FFT, quantitative research methods were assessed as suitable methods for the project.

The empirical work in this thesis is based on two types of data: Swedish official school administrative data and data from the IEA (International Association for the Evaluation of Educational Achievement) TIMSS-studies (Trends in International Mathematics Science Study). All studies within the thesis are based on different multivariate regression analysis methods. The different methods employed within each study are discussed below.

5.1. Study I

The aim of Study I was to examine the possible relationship between disruptive classroom behaviour and students’ outcomes at the school level in Swedish lower secondary schools. Since both these phenomena were expected to be related to school composition, the study also explored the relationship between these exogenous factors (such as student SES and immigrant background) and the endogenous factor of disruptive classroom behaviour.
5.1.1. Data

Data consisted of a joint dataset with official school administrative data from the Swedish National Agency for Education’s (Skolverket) ‘SALSA database’ for all Swedish lower secondary schools from the spring term of 2015 to the spring term of 2017 and matching data from the Swedish Schools Inspectorate’s (Skolinspektionen) ‘School Survey’. The dataset comprised variables on school composition, disruptive classroom behaviour and school results. In total, all 1 704 Swedish lower secondary schools, with their 102 105 9th grade students and 48 672 teachers were included in the gross sample. Due to missing cases, the net sample in the analysis was 1 007 schools, representing 59 percent of all lower secondary schools in Sweden.

5.1.2. Analytical strategy and methods

The research questions for Study I were answered through descriptive statistical analysis, including correlation analysis and multivariate regression analysis. Thus, the analysis could reveal potential relationships between school composition (exogenous factors) and disruptive behaviour (endogenous factors) as well as relationships between compositional effects and achievement. Since all data were at the school level, OLS multivariate regression analysis could be employed. All data preparation and analyses were conducted within SAS Enterprise Guide 7.15.

The main analysis was conducted in two steps using two different models. This was done to test the potential effects of disruptive behaviour beyond a traditional SES model for predicting schools’ results. The outcome of these analyses gave reason to conduct separate mediation analysis, which was done in accordance with the conditions set up by Baron & Kenny (1986). Sobel’s test for significant mediation was also conducted.

5.2. Studies II and III

Peer effects in education are a result of pupils having peers. Since most instruction is conducted in classes, the objects of interest for this research are students in classes. This is associated with some problems. First, students are very seldom randomly assigned to classes. This means that the methods
employed must have a multilevel approach to be able to handle this selection problem (Manski, 1993; Sund, 2009) and, secondly, the natural nested multilevel structure of school data (Heck, Thomas, & Tabata, 2010; Hox, 2010).

The shared aim of studies II and III was to test different versions of the FFT as a theory on peer effects. The main methodological alternatives were hierarchical linear modelling (HLM) and structural equation modelling (SEM). Both these methods are based on multivariate regression analysis. HLM is a more complex form of ordinary least squares (OLS) regression that can handle multilevel analysis. There are several examples of researchers employing this methods to study peer effects in education (Ma & Klinger, 2000) and opportunity to learn (Wang J., 1998). Thus, given the theory-testing approach of studies II and III, the framework of structural equation modelling (SEM) was assessed as a more suitable method.

5.2.1. Structural equation modelling

There were several reasons for choosing SEM as the analytical framework of these studies. First, SEM is a method that can be driven by theory to a large extent, which suits the research aims of the studies. Also, since many of the concepts to be investigated in these studies, such as OTL and the steering and limiting effects on teachers’ instruction, are not directly observable, the process of establishing measurement models within SEM using multiple items in confirmatory factor analysis (CFA) was assessed as a particularly useful approach.

The CFA approach is a theory-driven method. Within CFA, possible latent variables are derived from theory and then tested on the data. This requires a strong a priori sense of which latent variables should be derived and tested in the model (Kline, 2015). If the model does not fit the data, either a new model must be derived from the theory or the theory must be modified or rejected (Gustafsson J.-E., 2009; Kline, 2015).

Second, the SEM framework allows researchers to model complex relationships between both observed latent and manifest variables. Through path analysis, variables in multiple regressions can be specified as dependent
in one relationship and independent in a second (Muthén & Muthén, 1998-2017). In this sense, SEM offers a flexible ability to combine the handling of complex relations between dependent and independent variables, mediation analysis and analysis of latent variables (Kline, 2015). As discussed, according to the FFT, the teacher must be treated as both a dependent and independent factor when analysing classroom instruction, which SEM offers powerful ways of doing (Gustafsson J.-E., 2009; Wang, Hefetz, & Liberman, 2017).

Third, as the mentioned peer effects originate from between-classroom variation, an analysis of peer effects has to employ a multilevel approach to be able to handle both the selection problems in school systems (Manski, 1993; Sund, 2009) and the natural nested multilevel structure of school data (Heck, Thomas, & Tabata, 2010; Hox, 2010), which multilevel SEM analyses are able to do.

Beyond this, SEM also provides useful tools for ensuring construct validity and reliability due to its inherent ways of testing internal consistency between multiple indicators (Abedi & Baker, 1995; Abedi, 2002; Marsh & Hau, 2007; Zhao & Gallant, 2012). For these reasons and even more, SEM has grown rapidly within educational research (Gustafsson, 2009; Wang, Hefetz, & Liberman, 2017).

Taken together, these reasons led to the choice of applying multilevel structural equation modelling in studies II and III.

5.2.2. Assessing models within SEM

When assessing an SEM model, there are several different statistics – or fit indices – that can be used. The aim is to establish whether or not the derived model fits the data (Hu & Bentler, 1999; Kline, 2015). In each study, these indices are assessed and discussed.

One important statistic is the $\chi^2$ goodness-of-fit test used within CFA. The test offers a possibility to examine the deviation between the implicated (by the measurement model) and the actual observed covariance matrix in data. If the test returns a significant result ($p < .05$), this implies that there is a significant deviation between the implicated and observed matrixes and that the model
therefore does not fit the data (Kline, 2015; Wang, Hefetz, & Liberman, 2017). In studies on big samples, however, there is a problem with the \( \chi^2 \) test. The test almost always returns a significant value, even if the deviation between model and data is very small (Gustafsson J.-E., 2009). Thus, there are reasons to assess additional fit indices.

Another important statistic is the Root Mean Square Error of Approximation (RMSEA). The strength of RMSEA is that it measures the deviation between model and data on an absolute scale regarding sample size and the number of parameters in the model. The former means that the measure works well for studies with big sample sizes (such as TIMSS), and the latter means that it does not reward models with more parameters, which the \( \chi^2 \) test does (Gustafsson J.-E., 2009). The latter is important since adding more parameters is seldom theoretically relevant and therefore is also a violation of the principle within SEM that models should be as parsimonious as possible (Kline, 2015).

Two other important fit indices are the Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR). Both these measures are closely related to the RMSEA. The difference for SRMR is that it reports a standardized value on the deviations between model and data. The SRMR is also reported for each level in the multilevel analysis. CFI, on the other hand, is measured on a scale from 0 to 1, where 1 represents a perfect model fit. The measure can best be described as measuring how much better the implied model fits the data than does an empty baseline model.

One recurring strategy for assessing the results from the \( \chi^2 \) test in analyses on big samples is to assess the discussed alternative fit indices in relation to the results of the \( \chi^2 \) test (Eklöf & Knekta, 2017). If all other indices indicate a good model fit but the \( \chi^2 \) test does not, then it is probable that the large sample size has been a driver of the \( \chi^2 \) value (Gustafsson J.-E., 2009).

If a measurement model fits the data, then this will also be revealed by the share of variance the latent variables can explain in the manifest variables. For each manifest variable, the analysis will give an estimated effect of the latent variable. A non-standardized coefficient of such can be understood as an ordinary regression coefficient. Within the SEM framework, it is a common procedure to standardize the estimates. These estimates – or the ‘factor
loadings’ – can then be interpreted as Pearson’s correlation coefficients (Kline, 2015). According to Kline (2015), an ‘ideal result’ from the CFA analysis is that the measurement model explains at least 50 percent of the variance in each continuous predictor, meaning that each coefficient should be 0.70<. As said, this is an ‘ideal result’. Other researchers have claimed that estimates of 0.40< are acceptable (Wang, Hefetz, & Liberman, 2017), and there are examples of published studies with even lower loadings than that (Hansson, 2010). These examples highlight an important intersection between the need for statistical power, the risk of type 1 errors (rejecting a null hypothesis even though it is true) (Harring, Weiss, & Li, 2015) and the theory-driven approach of CFA analysis.

One final area within SEM that needs commenting is the possibility of handling mediation analysis in a flexible way. Within multivariate regression analysis, there are several procedures that can be implemented for handling mediation analysis (Baron & Kenny, 1986). Even though new techniques have been developed that facilitate analysis (Hayes & Matthes, 2009), they still require additional work. Within SEM, and with its powerful and flexible way of handling statistical modelling, the researcher has vast freedom to specify the relations within the model, making it easy to specify and test such effects.

Figure 8 illustrates a fictive SEM model with seven manifest variables, two latent variables and one dependent outcome variable. The figure also illustrates the symbols within SEM and how a mediation analysis (between LATENT 2, MANIFEST 7 and the outcome variable) is specified.

All modelling in studies II and III was conducted using Mplus 8.2 (Muthén & Muthén, 1998–2017). Maximum likelihood estimation with robust standard errors (MLR) was used for estimation. All data preparation and descriptive analyses were conducted in SPSS version 28.
Figure 8: A fictive SEM model

Note: Within SEM, manifest variables are marked with rectangles and latent variables with ellipses. A circle with ‘&’ marks the residual – the unexplained variance – for manifest variables. A single-headed arrow indicates dependence between an independent and dependent variable. A double-headed arrow indicates covariance. The figure illustrates a single-level model.

5.2.3. Data

Studies II and III are both based on Swedish TIMSS data. The decision to utilize this data was driven by the adequate population and sampling procedures within the studies, the level of data collection (including the classroom level) (Schmidt, Burroughs, Zoido, & Houang, 2015), the empirical aims of my studies and its availability. There are also several reasons to focus the empirical studies exclusively on mathematics. One reason is the hierarchical and cumulative nature of mathematics (Dumay & Dupriez, 2007; Rolfe, Strietholt, & Yang Hansen, 2021). From an FFT perspective, this is focal. Students’ prior knowledge in mathematics can in many cases be a necessary condition for the next domain to be taught. This can be exemplified by the relation between being able to calculate multiplication and division and being able to learn how to calculate percentages. The former is to a large extent
a necessary condition for the latter. This was also the reason why Lundgren (1972) focused many of his empirical studies on mathematics.

Another reason for choosing mathematics is that mathematics is one of the school subjects that is least influenced by home and background characteristics and most influenced by school processes, such as instruction (Rolfe, Strietholt, & Yang Hansen, 2021).

There are also both similar and somewhat other reasons to choose TIMSS over PISA. These include TIMSS’ focus on school learning, sampling of whole classes (Schmidt & Burroughs, 2013; Rolfe, Strietholt, & Yang Hansen, 2021) and possible validity issues with PISA’s OTL measures (Yang Hansen & Streitholt, 2018).

Study II was conducted using data from the Swedish TIMSS 2015. The dataset contained mathematics test results and survey data for 4,090 students in the 8th grade (on average 14 years old) and their 200 teachers in mathematics (Martin, Mullis, & Hooper, 2016; Skolverket, 2016; Foy, 2017). Data on the 8th grade was chosen for several reasons, including the fact that previous studies have shown that TIMSS test items have a stronger correlation to the Swedish 8th grade curricula than the 4th grade curricula (Skolverket, 2016).

Since the study was concerned with compositional effects in classroom instruction, there had to be more than a few students in the included classes. Only intact classes were included in the dataset. Descriptive statistical analysis revealed classes with missing data on class size, which were therefore excluded. Also, classes with less than 13 students proved to a large extent to be outliers in several variables included in the statistical modelling and test results in mathematics. These classes were therefore also excluded. The final dataset contained 3,761 students in 179 classes, representing 92 percent of the students and 87 percent of the classes in the original sample. The data were weighted using ‘house weights’ on the between-class level (Rutkowski, Gonzalez, Joncas, & von Davier, 2010).

The decision to also utilize TIMSS data for Study III sprang from the same arguments as for Study II, although Study III was conducted on data from the 2011 cycle, the same cycle used by Morton & Riegle-Crumb (2020). The reason for choosing the 2011 cycle was mainly to test a replica of Morton &
Riegle-Crumb’s specification of the ‘time on advanced content’ variable in the Swedish context.

The gross sample for Study III contained test results and survey data for 5,816 students in 8th grade (on average 14 years old) and their 364 teachers in mathematics. The gross sample contained classes that proved to be outliers in class size. Classes with less than 10 and more than 40 students were excluded. Also, classes with more than one math teacher were excluded because it would be impossible to decide which teacher’s assessment of the OTL variables (content coverage and time spent during the year) should be used in the models. The net sample after these exclusions was 3,928 students in 207 classes, representing 68 percent and 78 percent of the gross sample, respectively. Data was weighted using ‘house weights’ (Rutkowski, Gonzalez, Joncas, & von Davier, 2010).

5.2.4. Analytical strategy and methods

Both Study II and Study III were similarly conducted in different analytical steps. As mentioned, the first step in all SEM analyses is to establish the measurement model. My decisions on which variables to include in this process were based on my operationalization of the FFT, the FFT and OTL (Lundgren, 1972; 1973) and also on previous research within the field, specifically concerning OTL (Schmidt, Burroughs, Zoido, & Houang, 2015; Rolfe, Strietholt, & Yang Hansen, 2021) and measures of SES (Wiberg & Rolfsman, 2021).

In both studies II and III, there were reasons to test the full SEM models in different versions, continuously adding different variables. For Study II, this method concerned controlling the effects of adding different mediations in the analysis. For Study III, this method concerned exploring different steering and limiting compositional effects.

TIMSS employs a matrix-sample design that generates plausible values of student achievement in mathematics and science. Essentially, the plausible value method treats student achievement as a missing value. Responses to the limited test items and the student background questionnaires are then used in conjunction to generate a student ability distribution. This means that for each
student, there are five plausible values for mathematical achievement (BSMMATH01-05). To handle this in the analysis, the ‘imputation’ command in MPLUS was used. When doing this, MPLUS runs the model on each plausible value (1–5) and then reports the mean estimated effect and standard deviation between each estimation. The mean effect and standard deviation between each estimation were reported in both studies (Rutkowski et al, 2010).

5.3. Validity and reliability

5.3.1. Establishing validity and reliability

Assessing and establishing validity is one of the most important aspects of empirical social sciences. The problem originates from the fact that the issues we are investigating are formulated on the theoretical level, while the actual empirical work is done on the operational level. The core problem is how we operationalize our theoretical definitions (Esaiasson, Gilljam, Oscarsson, & Wängnerud, 2004).

If a study is valid, it means that the indicators of our theoretical definitions actually measure what we want them to measure (Bryman, 2008). This definition can be subdivided into two parts, construct validity and result validity. Construct validity is obtained when the theoretical definition and empirical operationalization correspond. Result validity is obtained when the measure actually measures what it is supposed to measure. Both these terms can be merged into the term internal validity. In contrast to internal validity, external validity concerns whether the obtained results can be generalized to the entire population of interest. When assessing the external validity, it is therefore of utmost importance to also assess the sampling strategy of the study (Esaiasson, Gilljam, Oscarsson, & Wängnerud, 2004).

There are no statistical methods to assess and establish validity. However, it is possible to create different operationalizations of the same theoretical definition and then test if they are correlated with each other. If they are, it can be an indication of construct validity but not definite evidence (Bryman, 2008).
In comparison, there are statistical methods of investigating reliability. A reliable measure is a measure with as few random errors of measurement as possible. Given this, a reliable measure should be possible to replicate in different samples and across time (Esaiasson, Gilljam, Oscarsson, & Wängnerud, 2004). Given this, a measure can be reliable without being valid, but a measure cannot be valid without being reliable (Bryman, 2008; Wikström, 2014). One common measure of reliability is Cronbach’s Alpha (Abedi & Baker, 1995). Cronbach’s Alpha stems from correlational analysis. The idea is to test the internal consistency among indicators that are believed to be indicators of the same theoretical definition. Just as Pearson’s R, Cronbach’s Alpha is reported on a scale between 0 and 1. In the literature, there are different recommendations for the required level, often between 0.60 and 0.80 (Esaiasson, Gilljam, Oscarsson, & Wängnerud, 2004; Bryman, 2008; Ahmad, Zulkurnain, & Khairushalimi, 2016).

As mentioned, one advantage of the SEM framework is its inherent ways of handling questions of validity and reliability. The theory-driven CFA analysis rests upon tacit assumptions on validity, and the CFA procedure also builds on the same concepts as Cronbach’s Alpha (correlational analysis), which means that fit indices within SEM that reveal an unacceptable model fit also indicate low reliability. In the literature, there are studies indicating that SEM could even be a better way to establish validity and reliability than conventional methods (Abedi & Baker, 1995; Abedi, 2002; Marsh & Hau, 2007; Zhao & Gallant, 2012), likely due to the fact that latent variables do not contain errors of measurement, which manifest variables do (Abedi & Baker, 1995; Gustafsson, 2009).

5.3.2. Validity and reliability of Swedish school administrative data

The Swedish school administrative data used for Study I was assessed as being valid and reliable. The statistics used for these measures are reported by the schools to Statistics Sweden (SCB), which employs rigorous controls of the reported data together with the Swedish National Agency for Education (Skolverket). Data on school composition is created through merges between the school-reported data and the total population register and register of education (for parents’ educational background) at Statistics Sweden. The
definitions for the data in Study I therefore are the same as those of Statistics Sweden, indicating good result validity and external validity. There are some missing cases in the data, but the rate of coverage enables generalizations for the entire population, demonstrating external validity for the study.

In contrast, the external validity is assessed as somewhat lower in the Swedish School Inspectorate’s (Skolinspektionen) School Survey since there are more missing cases (Skolinspektionen, 2019). Analyses within Study I revealed that the share of missing cases in the student surveys was higher in low SES schools than in high SES schools. The same pattern was not found concerning the teacher surveys. These problems of external validity were discussed in Study I.

When the school survey was established by the Swedish School Inspectorate, extensive inquiries were made concerning construct validity. Together with Statistics Sweden, cognitive interviews with students were conducted (Skolinspektionen, 2019). Descriptive analyses over time have shown that the measures of disruptive behaviour that are included in Study I are stable over time, indicating that the measure is reliable (Skolinspektionen, 2020). However, analysis of student and teacher assessments of disruptive behaviour in class revealed differences between the two groups (compare Kunter & Baumert, 2006). For instance, there were stronger correlations between teacher-assessed disruptive behaviour and student achievement than between student-assessed disruptive behaviour and achievement. This indicated that teacher-assessed disruptive behaviour was a stronger predictor of student achievement. This issue is further discussed in Study I since it indicates possible problems with the internal validity of the student questionnaires in the School Survey.

5.3.3. Validity and reliability of TIMSS

One of the main problems of validity when utilizing TIMSS data in a study is that the data must be considered as secondary data, i.e. that the data was not primarily collected for the study at hand (Smith E., 2008; Johnston, 2014). This means that I, as a researcher, have had no influence over the theoretical definitions, the operationalizations or the sampling strategy, and thus little influence over factors influencing both internal and external validity. From
this perspective, it is even more important to thoroughly examine the validity and reliability of the data.

**Does TIMSS measure mathematical competence?**

What 8th grade students ought to learn in school is strongly influenced by the national context in which they live. This affects how TIMSS results can be interpreted and used, and above all, how comparisons between different countries can be done (Lin, Bumgarner, & Chatterji, 2014). With this said, there is of course a vast overlap between different countries’ curricula in mathematics; many countries want their student to master the same type of mathematical skills. When constructing TIMSS, this has been a focal point – to establish which domains this concerns and how this could be measured (Wolf, 1998; Martin, Mullis, & Hooper, 2016).

The natural point of departure when assessing TIMSS’s result validity in the Swedish context is the Swedish curricula for 8th grade mathematics. For each TIMSS cycle, the Swedish National Agency for Education (Skolverket) conducts such analyses. These analyses have shown that the content of TIMSS is well-represented in the Swedish curricula, better so for the 8th grade curricula than for the 4th grade curricula. Of course, TIMSS does not contain all curricula content for Sweden, but the content that is represented in TIMSS is also found in Swedish curricula (Skolverket, 2016). This indicates good result validity of TIMSS’s test items in Sweden. Further analyses have also revealed a strong result validity since there is a strong correlation between Swedish students’ TIMSS results in mathematics and their subject grades and national test results (Skolverket, 2017; Wiberg, 2019).

Similar results have been found by Sollerman (2019). In his study, Sollerman investigated the relevance of the international large-scale assessments (ILSA) that Sweden takes part in and how these can be interpreted in a Swedish context. He describes his method as an ‘argumentative validation method’ in which the ILSAs are contrasted against Swedish curricula and instruction in mathematics. The analysis reveals that the content of the ILSAs, including TIMSS, is accommodated within Swedish mathematics curricula and context. Just as Skolverket (2016), Sollerman finds an ‘underrepresentation’ of Swedish content. The whole Swedish mathematics curriculum is not tested in PISA and TIMSS. Sollerman also verifies the results from Skolverket (2017)
that there is a strong correlation between Swedish students’ national results and their results in the ILSAs, indicating good result validity and external validity.

As understood by both Skolverket (2016) and Sollerman (2019), there are also test items in the ILSAs that are not represented in Swedish curricula or instruction. This can have a potential negative effect on student achievement in the tests. To investigate this effect, Test Curriculum Matching Analysis (TCMA) is conducted by IEA. The purpose of these tests is to investigate what effect it has on student test scores if such items – those not represented in the national curricula – are excluded (Wolf, 1998). Analysis of the Swedish case reveals very small effects of such (Skolverket, 2016).

In the Swedish debate on the declining results of the ILSAs, some have argued that this could be a result of students’ deteriorating motivation to write the tests since they are so-called low-stakes tests. (It does not matter to the students what results they attain since it will not affect their grades.) If this were true, it would be an indication of low reliability and low external validity for the ILSAs in the Swedish context. Eklöf & Knekt (2017) found in a study of Swedish data (including TIMSS) that Swedish students exhibit lower levels of motivation in taking ILSAs (low-stakes tests) than they do in taking other tests that will affect their final grades (high-stakes tests). Analyses from the Swedish National Agency for Education (Skolverket) (2015) have shown that this effect is negligible. The decline in Swedish results is explained by other factors, such as ‘the students’ do not have sufficient knowledge of the content that is being measured’ (p. 7, my translation). These results indicate a good internal and external validity of TIMSS in the Swedish setting.

Finally, another source for assessing the external validity of the ILSAs (including TIMSS) in the Swedish context is to compare the trends within different sources on students’ mathematical competencies. The Swedish National Agency for Education (Skolverket) summarized the trends in several ILSAs and found that they all exhibit the same trends (Skolverket, 2014). Similar results have been found in other longitudinal studies, such as diagnostic tests at technical universities in Sweden (Henrekson & Jävervall, 2016), indicating good internal and external validity and reliability in the ILSAs.
Surveys within TIMSS

Alongside the subject-related tests, both students and teachers answer background surveys in TIMSS. From these, the different indicators are derived. One example concerns students’ motivation in taking the tests, which Eklöf & Knekta (2017) utilized in their study.

Just as with the curricula, there are strong contextual factors affecting how respondents will respond to the same survey items. In a study of TIMSS measures of student motivation in Anglo-Saxon and Arab countries, Marsh et al. (2013) found both within- and between-group variance in student motivation that was related to contextual factors. Some of this variance could be explained by how words such as negations were used in the survey items and also by how schools were organized (mainly whether or not they were co-ed). Despite this, Marsh et al. could conclude that the correlation between motivation and achievement was very similar across countries, indicating a general internal validity and reliability of TIMSS.

Eklöf & Knekta (2017) found similar results in their study. Referring to previous international studies, they found an average correlation of $r = 0.30$ between motivation and achievement. In comparison, they reported correlations of $r = 0.25–0.48$ between motivation and achievement in PISA, TIMSS and regular Swedish national tests. Smith & Lopez-Castro (2017) reported similar results when comparing PISA and TIMSS across countries, further strengthening the case for the internal and external validity and reliability of the surveys.

In the Swedish context, there are further specific indications of good validity and reliability of TIMSS. With a cycle of three years, the Swedish National Agency for Education (Skolverket) since the 1993/94 semesters have investigated ‘Attitudes toward the school’, a big national representative survey where students and teachers answer questions about school organization and their attitudes towards it. In the most recent study from 2018 (2021 was cancelled due to the Covid-19 pandemic), a similar decrease in students’ desire to learn in school was found, just as the TIMSS survey did (Skolverket, 2019), indicating both the internal and external validity and reliability of TIMSS in the Swedish setting.
There is one area where TIMSS shares the same validity problems as other surveys of youth and students. That is questions that concerns their parents’ occupational status, educational background, and such (information that is then used to establish students’ SES). Questions like these tend to be hard for students to answer, especially when they are younger.

Engzell has made several studies on the issue. In one study, Engzell & Jonsson (2015) estimated the effects of such errors of measurement when predicting students’ cognitive abilities. By merging survey data from the large-scale study of ‘Children of Immigrants Longitudinal Survey in Four European Countries’ with Swedish official register data, they could investigate the differences between the responses they gave about their parents and the actual information that could be found in the registers. They found no major errors in how students reported their parents’ occupational status but more errors when they reported on their educational background. The reported information tends to show less variance than there actually was, which could lead to underestimating the effects of parental educational background.

Concerning the same issue, one popular indicator in sociological research on children is to ask how many books there are in their homes. The question is supposed to be an indicator of cultural capital. In a study on data from PIRLS, Engzell (2019) investigated the validity of the indicator. By comparing how different students answer the question (for instance boys and girls) but also how parents answer the same question, he found that low-achieving students tend to underestimate the number of books there is in their home. Engzell then demonstrates what kind of effects this has on reported results and analyses of differences between countries. Engzell’s findings are important, but in reference to TIMSS 8th grade surveys, it is important to bear in mind that the students are older than those of PIRLS (4th grade), which likely will be associated with fewer measurement errors.

In a similar study, Wiberg & Rolfsman (2021) predicted differences in using self-reported SES indicators in TIMSS with register-based information from Statistics Sweden (SCB). They found that, if possible, one should use register-based data. When this is not possible, the TIMSS Home Educational Resource index (HER) or number of books at home is preferable. These indicators ‘gave similar correlations with the achievement measures as the register-based
variable of parents’ highest education’ (p. 9). The recommendations of Wiberg & Rolfsman were applied in studies II and III.

Concerning teacher-reported data, above all in reference to different OTL measures, there are reasons to discuss validity issues. One concern could be related to possible teacher bias when reporting on characteristics of their own instruction, teaching practices and quality due to possible self-serving strategies or teaching ideals. In a study on these issues utilizing PISA data, Kunter & Baumert (2006) found specific conceptual structures for student and teacher ratings where teachers elaborated on the use of tasks and methods in instruction and students focused on the support from their teacher in different aspects. They found three shared dimensions between students and teachers. These were occurrences of classroom management problems, degree of cognitive autonomy during lessons, and tempo of interactions. They concluded that student and teacher ratings are best suited for capturing different aspects of the learning environment and instruction.

On the same theme, Porter (2002) summarized previous research into validity issues of survey data on OTL, mainly in mathematics. He concluded that

‘[g]enerally, these investigations find that survey data is excellent for describing quantity – for example, what content is taught and for how long – but not as good for describing quality – for example, how well particular content is taught. (p. 9)

In reference to these results, studies II and III both utilize students’ assessments within the OTL dimension of instructional quality, including assessments of the type of support their teachers provide in instruction. Several indicators from the TIMSS construct of ‘Instructional clarity’ were used (Fishbein, Foy, & Yin, 2021). Concerning the OTL dimension of content coverage, on the other hand, teachers’ assessments were used. The methods employed by IEA in TIMSS – to ask teachers whether or not their students have been taught a certain material mostly before this year, this year or not yet taught – have proved to be reliable indicators of content coverage (Wang J., 1998; Porter, 2002; Rolfe, Strietholt, & Yang Hansen, 2021).
5.3.4. **Concluding remarks on validity and reliability**

There are several indices of the validity and reliability of the data used in the studies in this thesis. The school administrative data for Study I is defined and collected by Statistics Sweden, warranting the result validity as well as the external validity of the study. However, some issues concerning the differences in result validity of the Swedish School Inspectorates’ School Survey, between students’ and teachers’ assessment of disruptive classroom behaviour (Kunter & Baumert, 2006), were discussed. Concerning the external validity of Study I, the effects of missing cases were investigated and discussed but revealed no signs of extensive problems with validity.

Concerning studies II and III and the use of TIMSS data, several issues of validity and reliability were discussed. One issue concerns the concept of SES. Previous research displayed several validity issues concerning students’ self-reported data on parents’ educational background and occupational status (Engzell & Jonsson, 2015; Engzell, 2019). In the Swedish context, Wiberg & Rolfsman (2021) conducted thorough analyses of the validity of TIMSS SES measures related to Swedish register data on student background and achievement. Their recommendations have been taken into account in studies II and III.

Another issue concerns the concept of OTL and the measurement of its different dimensions (Stevens, 1993; Porter, 2002). Here, the validity and reliability of teachers’ self-reported assessments of their instruction were discussed. Following similar conclusions as those of Porter (2002) and Kunter & Baumert (2006), studies II and III utilized student-assessed indicators for the measurement of the quality of instructional delivery and teacher-assessed indicators for the measurement of content coverage (Wang J., 1998).

One overall issue for this thesis concerns causality. In the ‘language of SEM’, the term ‘effect’ is frequently used (Kline, 2015), implying that the relations that are being investigated concern causal effects. This is not the case in this thesis. Especially concerning studies II and III, the cross-sectional design of TIMSS does not enable causal inference, only correlational analysis between variables. Despite this, the theoretical framework concerns causation. This translates into a problem of external validity for the thesis.
Some researchers have argued that ILSAs, under certain circumstances and with certain analytical strategies and methods, can be used for causal inference (Gustafsson & Nilsen, 2022). None of these methods have been used in this thesis – why causal inference is impossible from a methodological perspective. In relation to the theoretical perspectives of the thesis and some practical circumstances concerning the data utilized, causal reasoning about the results could still be plausible. The main argument for this concerns the time order of the independent and dependent variables in the analysis. There is a clear directionality from the independent to the dependent variables (Rolfe V., 2021). One example of this is the OTL construct of content coverage, where teachers indicate when in time instruction concerning specific parts of the mathematical curriculum were taught (before the test). Another example concerns the class composition, which also precedes students taking the test in TIMSS. For Study I, a similar argument could be applied since the School Survey and the measurement of disruptive classroom behaviour for all schools precede the final grades for the students.

5.4. Research ethics

All data used in this thesis – ranging from Study I to Study III – are publicly available and provided fully anonymized. The data for Study I was retrieved from the web pages of the Swedish National Agency for Education⁴ (Skolverket) and the Swedish School Inspectorate⁵ (Skolinspektionen). The data and documentation for Study II and III were accessed via the IEA’s online

⁴ https://www.skolverket.se/skolutveckling/statistik
⁵ https://www.skolinspektionen.se/beslut-rapporter-statistik/statistik/statistik-fran-skolenkaten/
data repository\textsuperscript{6} with valuable help from the Swedish National Agency for Education (Skolverket). Due to the public and anonymized nature of data in this thesis, authorization for this project was not sought from Jönköping University’s ethics committee.

\textsuperscript{6} http://www.iea.nl/data-tools/repository
6. Empirical findings and integrated discussion of original studies

The aim of this thesis is to make a theoretical contribution to the field of peer effects and compositional effects in education. The thesis orbits around one main argument as described in the introduction:

(i) The Frame Factor Theory (FFT) can be adopted as a theory on peer effects in education.

(ii) The FFT predicts that classroom practice, content coverage and study environment are somewhat dependent on class composition. These factors are mechanisms that generate peer effects and compositional effects on student achievement.

(iii) Since Opportunity to Learn (OTL) investigates differences in educational provision, for instance through content coverage, there is a theoretical overlap between OTL and FFT.

(iv) By premises (i–iii): A theoretical synthesis of the FFT and OTL would hypothetically create a suitable theory with predictive power on peer effects and compositional effects within classroom instruction.

As mentioned in the introduction, the argument is a deductive argument. This means that the argument can only be true if and only if the premises are true. The thesis consists of three empirical studies that each correspond to the different premises (i–iv) in the argument. In this chapter, each study will be first summarized and discussed in reference to the premises and, secondly, they will be considered in an integrated discussion to try to establish whether the main argument is true or not. Finally, the chapter will discuss the contribution of the thesis to the field of peer effects in education and suggest future research.
6.1. Study I: School Composition, Disruptive Classroom Behaviour and Student Results: A Study of Mechanisms of Peer Effects

The aim of the study was to examine the possible relationships between disruptive classroom behaviour, school composition and students’ outcomes in the Swedish setting (premise ii), following previous studies by Lavy & Schlosser (2011) and Lavy, Paserman, & Schlosser (2011), with the intention of ‘setting the stage for a planned future revisit to the frame factor theory’. Through this, the study contrasts previous research, which often focuses on the immediate peer effect on individual students’ achievement, not how possible peer effects arise from the interaction between class composition, pupil-pupil interactions and teacher-pupil interactions. In contrast, the study focuses on the interaction between the group (the class) and the teaching process, as predicted by the FFT.

The study verified a significant relationship between a school’s pupil composition and disruptive behaviour. The higher the average level of parental educational background at school, the better the teachers assessed the classroom environment with less disruptive behaviour. As expected from previous research, a higher share of newly arrived immigrants was related to more disruptive behaviour.

The results also revealed that disruptive classroom behaviour was related to the schools’ results measured as their mean GPA, in part confirming previous results of Lavy & Schlosser (2011), Lavy, Paserman, & Schlosser (2011) and also those of Cheema & Kitsantas (2014) and Grøne Kristoffersen et al. (2015).

Considering the distinction made between endogenous and exogenous factors in the economic literature on peer effects, the results of this study provide interesting input. The analysis indicates that some of the original effects of schools’ pupil composition were in fact mediated through disruptive classroom behaviour. This means that some of the negative effects that, for instance, the share of boys has on schools’ results, was in fact an effect of disruptive classroom behaviour. Disruptive classroom behaviour is thereby a limiting factor in the teaching process, as predicted by the FFT.
In reference to the main argument of the thesis and the more precise premise (ii), the results of the study support the view from previous research on the ‘intertwined nature’ of schools’ functioning (Barr & Dreeben, 1983; Dreeben & Barr, 1988; Thrupp, 1999). Compositional effects seem to operate through a nested series of hierarchical layers, where school, class and group compositional effects are intertwined (Garrison, 2004). Due to the study design and available data, it is important to note that Study I does not support the full width of premise (ii). It especially lacked the relation to OTL, which was developed in Study III with guidance from the results from Wang, Liu, & Leung (2022).


The aim of the study was to test the hypothesis that the FFT can be adopted as a theory on peer effects in education, as it was formulated in premise (i) of the thesis’s main argument. From the FFT, a multi-level model of classroom instruction was derived, operationalized and tested empirically on Swedish TIMSS data of 2015. The hypothesis was that between-class variance in student achievement to some extent must be explained by the FFT construct of steering and limiting effects, which would be related to the level of the steering criterion group (SCG) in each class.

The study revealed a strong relation between the FFT construct of steering and limiting effects on teachers’ instruction and student achievement. Under control of the class SES, students in classes where teachers reported lower levels of steering and limiting effects on the instruction scored significantly higher results in TIMSS mathematics. Thus, the study gave empirical support to premise (i) in the main argument.

As it turned out, the study could not verify the modus operandi of the FFT construct of steering and limiting effects on teachers’ instruction. According to the FFT, the level of the SCG will affect the content, timing and pacing of instruction. Lundgren (1972) operationalized this phenomenon in curricula units covered, their level of complexity, and time spent. Following the results
of Hansson (2010; 2012), two latent variables were specified in the study to try to capture the different dimensions of teachers’ responsibilities in instruction. According to Hansson’s results, the dimensions of teacher responsibilities differ between classrooms due to their composition.

The first latent variable was associated with teachers’ responsibilities for covering elementary parts of the curricula, and the second latent variable with more advanced parts. The two latent variables were not correlated, indicating that they represented two different dimensions of teacher responsibilities. Yet, the results showed no relation to limitations of instruction or student achievement, as predicted by the FFT. Even though the measurement model showed a good fit to the data, it seemed impossible to rule out the possibility of poor construct validity for these two latent variables in reference to the FFT. Hereby, the study summoned future research into peer effects and the FFT to further investigate whether such effects could be verified, turning the theoretical interest towards the OTL perspective on instructional provision.


Since Study II could not verify the modus operandi of the peer effects, it was discussed if this could be due to poor construct validity of the instruction-related variables in the model. Together with the other empirical findings of Study II, the results highlighted the need for a theoretical synthesis between OTL and the FFT, as formulated in premises (iii) and (iv) in the main argument of the thesis. A synthesis of such would also bridge the distance between research into educational inequality, educational opportunities and peer effects in education.

The aim of Study III was to further develop the theoretical framework used in Study II and test a revised model on peer effects in Swedish schools. The model had to include a combination of the FFT prediction of class compositional steering and limiting effects on teachers’ instruction and the OTL perspective of inequalities in educational provision, mainly operationalized as content coverage. From the synthesized model, an
operationalized model was derived and tested empirically on Swedish TIMSS data of 2011.7

The results verified that class composition was related to the FFT construct of steering and limiting effects on instruction and advanced content coverage, both of which were related to student achievement. The study thereby indicated that class composition to some extent was related to teachers’ instruction in class, through its limitations on instruction but also through the advanced content that can be covered in instruction. Teachers in classrooms where students’ prior knowledge and aptitude on average could be expected to be higher could to some degree cover more advanced content in instruction than teachers in other classrooms. The more and older the immigrant students were when arriving in Sweden, the less advanced was the content covered in class (Szulkin & Jonsson, 2007; Brunello & Rocco, 2013).

In the full model (M3), there was a positive relation between class SES on the between-class level and advanced content covered. There was also a strong robust negative relation between the FFT construct of steering and limiting effects on instruction and student achievement, indicating the presence of a peer effect. In the full model, there were no significant relations between the FFT construct of steering and limiting effects on instruction and the OTL constructs of advanced content coverage and quality of instruction.

Throughout the models in Study III, there was no relation between the FFT construct of steering and limiting effects on instruction and advanced content coverage, indicating that these are two separate phenomena, at least within the

7 The TIMSS 2011 cycle was chosen to be able to perform an exact replica of Morton & Riegle-Crumb’s (2020) ‘time spent on advanced content’ variable since it was assessed as theoretically relevant from the FFT perspective. The results of the replication are discussed in Section 6.4 below.
models. The predicted relations instead occurred with class composition. These results summon future research to test the predictions of the synthesized theoretical model in a longitudinal setting, with data on students’ actual prior knowledge in mathematics, just as the FFT was originally formulated and tested.

The results from the study seem to support previous findings on the steering and limiting effect of class composition on teachers’ instruction as well as providing a new theoretical perspective on results from OTL research (Stevens, 1993; Kurz A., 2011; Elliot & Bartlett, 2016; Klapp & Jönsson, 2020; Morton & Riegle-Crumb, 2020). Differences in students’ opportunities to learn seem to stem from, on the one hand, content coverage, and on the other hand, the steering and limiting effect that class composition has upon teachers’ instruction.

6.4. Integrated discussion

6.4.1. Summarizing and interpreting the empirical findings

The three studies, I, II and III, when taken as a whole, provide indicial evidence for the argument in the thesis. The results from Study I indicate that classroom management and instructional practices are related to class composition (premise ii), just as previous research has indicated (Barr & Dreeben, 1977; Bourke, 1986; Weinert, Schrader, & Helmke, 1989; Thrupp, 1999; Lazear, 2001; Garrison, 2004; Aizer, 2009; Lavy & Schlosser, 2011; Bertrand & Pan, 2013; Cheema & Kitsantas, 2014; Grøne Kristoffersen et al, 2015). From a classroom managerial perspective, the theoretical understanding of the empirical results of Study I is important. Recalling Lundgren (1972), the FFT predicts that different classes due to their composition will need different amounts of instruction time for learning specific curricula units. The results of Study I indicate that, in addition to a group’s academic prerequisites, determining the time needed to achieve curricula goals, the classroom environment also seems to be important to consider as this seems to be one of several other factors generating peer effects in education.
Regarding this issue from a teacher’s perspective enhances the interpretation. Having to spend a lot of time in class maintaining order and ensuring that the classroom environment is appropriate steals time from instruction and generates time-off tasks for students (Arlin, 1979; Anderson L., 1981). Thus, disruptive classroom behaviour contributes to creating a bigger gap between the instruction time needed by the class and the actual time spent on instruction. Disruptive classroom behaviour is thereby a limiting factor in the teaching process, as predicted by the FFT (Lundgren, 1972). The validity of this interpretation seems enhanced through Study I since this was measured in one of the indexes with the survey question ‘Maintaining order during class takes a lot of time from instruction’.

In the same manner, the FFT construct of steering and limiting effects on instruction in studies II and III contains one teacher survey item, as follows: ‘In your view, to what extent do the following limit how you teach this class? – Disruptive students’. The construct of ‘limitations on instruction’ in studies II and III is the main operationalization of the FFT prediction of compositional steering and limiting effects on instruction. Since both Study II and III provide evidence that this construct is a strong predictor of mathematical achievement on the between-class level, the results support the hypothesis that the FFT can be adopted as a theory on peer effects in education (premise i).

The theoretical interpretation of the FFT in studies II and III is that the compositional steering and limiting effects on instruction is a mechanism generating peer effects. Still, this does not reveal the modus operandi of this mechanism. The mechanism according to the FFT is that the level of the SCG will affect the content, timing and pacing of instruction. Lundgren (1972) operationalized this phenomenon in curricula units covered, their level of complexity and time spent. In OTL terminology, this translates to the ‘enacted curriculum’.

In Study II, the available data on time spent in TIMSS 2015 was assessed as inadequate. Therefore, following Hansson (2010; 2012), two latent variables measuring teacher responsibility for learning were specified using variables from the teacher survey. The first latent variable was specified to measure the presence of ‘teacher responsibility for content’, which was intended to capture
responsibility for elementary parts of the curricula. The second was specified to measure the presence of more ‘Advanced Instruction’.

Even though the measurement model showed a good fit to the data and that there was no relation between the two latent variables (indicating that they were two separate phenomena), the structural equation model revealed no relations or effects of the two as predicted by the FFT. This spurred a discussion in Study II about whether this could be due to poor construct validity for the two latent variables in reference to the FFT.

Referring to this discussion, Study III argues that there is a need for a theoretical synthesis of the FFT and OTL (premises iii–iv). There is an apparent theoretical overlap between the two in the perspectives on foremost content coverage in instruction. The study utilized TIMSS 2011 data to test the new synthesized model. Among other reasons, this was also done to try to replicate the ‘time spent on advanced content’ that Morton & Riegle-Crumb (2020) specified in their study on OTL in the US.

As previous research has shown, there is a very small between class/school variance in OTL in Sweden measured as content coverage in both PISA and TIMSS (Schmidt, Burroughs, Zoido, & Houang, 2015; Rolfe, Strietholt, & Yang Hansen, 2021). In reference to this, the empirical findings on OTL in Study III must be understood. The study revealed that a replica of Morton & Riegle-Crumb’s (2020) ‘time on advanced content’ construct was inappropriate for the Swedish context. Morton & Riegle-Crumb (2020) found that in US schools that were predominantly black, teachers on average spent 72 percent of class time during the year covering algebra and advanced content, while the corresponding share in predominantly non-minority schools was 82 percent. The analysis of the Swedish data revealed, first, that Swedish teachers in 8th grade mathematics on average spent much less time on algebra and advanced content than their colleagues in the US. Morton & Riegle-Crumb (2020) reported a mean of 81.6 (SD = 13.6%) while Swedish teachers only reported a mean of 41.8 percent (SD = 10.4%). Secondly, in the Swedish case, there were almost no significant correlations between time spent on advanced content and class composition. The correlation between time spent on advanced content and class SES was $r = 0.06 (p < .01)$ and age of coming
0.03 (p < .10). Thirdly, there was no correlation between time spent on advanced content and the class mean results (r = 0.02, p = .27).

Instead, Study III utilized a derived construct from TIMSS's 19 different content coverage variables, created through correlational analysis and the assessments of five Swedish 8th grade teachers in mathematics. The derived variable, ‘advanced content’, had a small positive effect on individual students’ achievement. In reference to the main argument in the thesis and the establishment of whether it is true or not, the explanatory power added by the addition of the OTL construct of advanced content must be evaluated and discussed. As Study III revealed, a model specification – excluding the OTL variables of advanced content coverage – explained 66.2 percent of the between-class variance in achievement. With OTL content coverage added, the model explained 68.7 percent of the variance.

Regarding the main argument of the thesis, and especially premises (iii) and (iv), the results from Study III seem somewhat inconclusive. The study verified the predicted relations of class composition and steering and limiting effects on the one hand and class composition and content coverage on the other hand. But in the full model, there are no mediating effects between class composition and content coverage through the FFT construct of steering and limiting effects on instruction. However, this was the construct that showed the strongest relation to student achievement. In reference to the main argument of the thesis, these results must be interpreted as indicial evidence for premises (iii) and (iv), indicating that further research is needed.

**6.4.2. Future research**

As discussed, the results of Study III indicate that there is a relation between class composition and advanced content coverage but not a specific relation – in this case – between ‘limitations on instruction’ and advanced content coverage. This would promote an interpretation that ‘limitations on instruction’ is not a mechanism generating that sort of peer effect that would be mediated through advanced content coverage.

Yet again, there is reason to return to a discussion on construct validity in reference to the FFT. As described in Chapter 4, the FFT rests upon crucial
concepts introduced by Carroll (1963). The main determinant of ‘time needed to learn’ according to Carroll is students’ prior knowledge. In the latent variable of ‘limitations on instruction’ in studies II and III, this construct is included together with other constructs, such as disruptive students. At the same time, due to its cross-sectional design, TIMSS lacks measures of students’ actual prior knowledge. This indicates that one important issue for future research to investigate, with data on students’ prior knowledge, is whether there is such a relation between class composition, steering and limiting effects on instruction and content coverage, as the synthesized FFT model in this thesis predicts.

Also, as discussed, the OTL gaps in Sweden are very small (Schmidt, Burroughs, Zoido, & Houang, 2015; Rolfe, Strietholt, & Yang Hansen, 2021), which most likely is related to the Swedish comprehensive model of schooling in reference to stratified systems (compare Dollmann & Rudolphi, 2020). The findings also seem to be well in line with previous findings concerning OTL measures in TIMSS in general (Flodén, 2002), which probably also affects the type and sizes of effects found in the data. This calls for future research to test the synthesized FFT model in this thesis in different countries and educational settings where (for instance) OTL gaps are bigger. Since the findings reported in this thesis are average effects for all students in class, future research is also summoned to explore possible interaction effects between compositional effects on teachers’ instructional practices and different students’ achievement.

6.4.3. In closure

The contribution of this thesis to the field of peer effects in education is mainly theoretical. The thesis has shown that the Frame Factor Theory is a suitable theoretical framework for the study of peer effects. Also, there is an important theoretical overlap between the FFT and Opportunity to Learn. The FFT predicts that teachers will be able to cover different content, at a different pace and in different classes due to students’ prior knowledge. In this sense, the FFT revives and injects some previous theoretical perspectives into future OTL research.
Within some OTL research, it is possible to detect a line of reasoning about teachers’ instruction that holds a view that the teacher consciously makes decisions, for instance, on what content to cover and in what manner for different classes, generating different opportunities to learn for different students (Wang J., 1998; Porter, 2002; Klapp & Jönsson, 2020; Tengberg et al., 2022). This line of reasoning is related to the ‘self-fulfilling prophecy’ perspective of Rosenthal & Jacobson (1968), namely that teachers design their instruction according to their expectations from their students, which delimits students’ possible learning and achievements.

This perspective was questioned by Lundgren in the formulation of the FFT. Lundgren wrote:

Why have different theories on teaching had so little influence on the teaching process? The answer seems to lie in the fact that these theories are not based on empirical analysis of teaching. Instead, they have built on different ideas of what teaching is. One such idea is that the teacher consciously determines the teaching. We hold that his steering can be done only within the limits set by the teaching frames. (Lundgren, 1972, p. 38f)

Several other researchers have shared Lundgren’s perspective (Barr & Dreeben, 1983; Thrupp, 1999) or as put by Weinert, Schrader, & Helmke:

[...] the effects of teacher behaviors tend to be more or less confounded with context factors. Moreover, some context factors (particularly, [students’] level of prior knowledge and ability) are not only determinants of teacher behavior, but also regarded as aptitudes determining subsequent achievement outcomes. Student characteristics, such as prior knowledge and ability, obviously can have a direct as well as an indirect impact (via instructional processes) on student achievement. (1989, p. 907)

The result from this thesis hereby summons future research into peer effects and OTL to view the teacher as not just an independent variable – consciously making decisions on the forthcoming instruction and enacted curricula – but also a dependent variable, who to some extent is steered and limited by the factors framing the classroom instruction.

From a wider perspective, the findings reported in this thesis are also interesting from a historical and political point of view. In Sweden, the predominant aim of educational policy throughout the post-war era has been
to increase school equity, mainly by decreasing educational differentiation through less streaming and tracking by students’ backgrounds and abilities (Beach, 2018). This policy has over time created a school system that from an international perspective has a heterogenous school composition regarding students’ background, small between-school OTL gaps, and with highly equitable results (Schmidt, Burroughs, Zoido, & Houang, 2015; Rolfe, Strietholt, & Yang Hansen, 2021).

From this perspective, the findings in this thesis can be seen as both promising and discouraging. The findings are promising in the sense that the compositional effects are not greater. To some extent, the Swedish comprehensive school model seems to work. But at the same time, the findings are discouraging since they reveal compositional effects on content coverage and student achievement of the same nature as those reported by Dahllöf in the 1960s (Dahllöf, 1967). Despite the aim of Swedish educational policies, it still matters – as it did in the 1960s – who students are, what background they have and who their classmates are.

In this sense, the thesis also provides important input to the contemporary Swedish debate on school segregation and increased between-school variation. History shows that the Swedish model of comprehensive schooling in heterogenous classes led to small OTL- and achievement gaps between schools. As a result of increased school segregation, the opposite development now seems to be occurring (Gustafsson & Yang Hansen, 2018; Yang Hansen & Gustafsson, 2019; Rolfe, Yang Hansen, & Streitholt, 2022). In reference to the results in this thesis, it seems unrealistic to expect that teachers in different schools would be able to halt this development just by their instructional practices. Teachers can make a difference, but not that big of a difference. Ultimately, teachers to some extent are steered and limited by the conditions given to them by the frame factors.

The collective nature of classroom teaching (Nuthall, 2005), or ‘the negotiated nature of teaching’ (Thrupp, 1999, p. 157), might force us to be somewhat realistic about the extent to which educational inequalities can possibly be eliminated in school learning.
References


83


Exploring Mechanisms of Peer Effects in Education
– Frame Factor Analyses of Classroom Instruction

Despite a vast international literature, previous research into compositional effects and peer effects in education has remarked on the lack of consensus within the field. To date, peer effects in education and students’ different opportunities to learn in school have been studied separately. In this thesis, it is argued that these perspectives need to be synthesized.

In the thesis, Frame Factor Theory is employed as a theory of peer effects in classroom instruction. According to the theory, one mechanism generating peer effects is the steering and limiting effect that class composition has on teachers’ instruction. It is argued that this perspective also needs to be synthesized with the Opportunity to Learn-perspective, to widen our understanding of why different students meet different instruction and thus are given different opportunities to learn. The theoretical perspectives and empirical models are evolved through three empirical studies included in the thesis. The theoretical development and the empirical results from each study is discussed in an integrative essay in the thesis.

The results indicate that class composition affects the presence of limitations on instruction and content coverage in instruction, both of which is related to students’ opportunities to learn and individual students’ achievement. Implications for future research and educational policy is discussed.

PONTUS BÄCKSTRÖM is a certified teacher in social sciences and philosophy. A part of a career as a secondary school teacher, he has worked for the Swedish School Inspectorate and the Swedish Ministry of Education. Parallel to being a PHD-candidate, he works for the Swedish Teachers’ Union.