

Teachability and School Achievement
Is Student Temperament Associated
with School Grades?

– Don't stop thinking about tomorrow –
To my son Ilari and daughter Elina

Sari Mullola

Teachability and School Achievement

Is Student Temperament Associated with
School Grades?

*Academic Dissertation to be publicly discussed, by due
permission of the Faculty of Behavioural Sciences at the
University of Helsinki main building, Unioninkatu 34,
Auditorium XII, Friday 30th November 2012, 12 o'clock*

*Supervisors: Professor
Liisa Keltikangas-Järvinen
University of Helsinki*

*Docent
Mirka Hintsanen
University of Helsinki*

*Professor
Niklas Ravaja
University of Helsinki*

*Pre-examiners: Academy Professor
Erno Lehtinen
University of Turku*

*Professor
Päivi Atjonen
University of Eastern Finland*

*Custos: Professor
Liisa Keltikangas-Järvinen
University of Helsinki*

*Opponent: Professor
Mare Leino
Tallinn University, Estonia*

ISBN 978-952-10-7865-1 (nid)

ISBN 978-952-10-7866-8 (pdf)

ISSN 1799-2508

Unigrafia

2012

Sari Mullola

Teachability and School Achievement—Is Student Temperament Associated with School Grades?

Abstract

School achievement has far-reaching consequences for the educational pathways and later life trajectories of youth. It determines a student's success or failure in being eligible for further studies, and consequently, career opportunities later on. Student temperament (i.e., an individual's unique and innate tendency to approach and react to the environment) and teachability (i.e., a teacher's perception of the attributes of an ideal model student) have been shown to be influential factors in predicting school success and academic outcomes measured by both standardized achievement tests and teacher-rated school grades. Consequently, student temperament has also been associated with student-teacher interactions, teacher expectations and educational decisions regarding a given student.

The purpose of the study was to examine how teachers perceive the innate temperament of students in the school context and whether teachers' perceptions of student temperament are associated with their perceptions of students' educational competence (EC) (i.e., cognitive ability, motivation and maturity) and teachability. In addition, the study aimed to determine whether teacher-perceived temperament, EC and teachability are related to student school achievement in terms of teacher-assigned school grades in Mathematics (Math) and Mother Language (ML). Teacher and student gender and teacher age were examined as moderating variables in the association between teacher-perceived temperament and school achievement.

The subjects were Finnish ninth grade adolescents ($n = 3212$) and their Math and ML teachers ($n = 221$) derived from a population-based sample of Finnish upper-comprehensive schools. Teacher-rated student temperament was assessed using age-appropriately formed scales from the Temperament Assessment Battery for Children—Revised (TABC-R) and the Revised Dimensions of Temperament Survey (DOTS-R). EC was assessed with three subscales covering cognitive ability, motivation, and maturity developed for the current study. Both teacher-rated and self-rated student teachability was assessed using the results of the factor analysis examined for the current data. The respective school grades were taken from the students' latest school reports for Math and ML.

The main findings were as follows: (1) teacher-perceived student innate temperament, EC, and teachability explained a rather high proportion of teacher-assigned student school grades; (2) teachers' ratings seemed to vary systematically by their gender, age and student gender; (3) the variance between teacher-rated and self-rated temperament could be explained by teacher-ratings and self-ratings measuring different concepts and explaining different variance; and (4) on the basis of findings 1–3, teachers' perceptions of student innate temperament could bias their assigning of grades.

The results suggest a need for more consideration regarding the prevailing assessment practice in the Finnish educational system, where the student learning process and lesson activity are central. Furthermore, the findings suggest that future teacher training should take into account temperament-conscious education.

Keywords: temperament, school achievement, teacher-rated, teachability, educational competence

Sari Mullola

Mallioppilasodotukset ja koulumenestys – onko oppilaan temperamentti yhteydessä kouluarvosanoihin?

Tiivistelmä

Kouluarvosanoilla on merkittävä ja pitkäaikainen vaikutus oppilaan tulevaisuuteen mm. jatko-opintoihin pääsemisen ja tulevan uravalinnan kannalta. Arvosanat eivät kuitenkaan perustu pelkästään standardoituin koulusaavutustesteihin, vaan sisältävät opettajan käsityksiä oppilaan taidoista ja taidoista heijastaen opettajan mallioppilasodotuksia, mielipiteitä, arvoja ja asenteita. Aikaisempien tutkimusten mukaan oppilaan temperamentti (yksilön synnynnäinen, yksilöllinen taipumus lähestyä uusia asioita ja reagoida ympäristöön) on yhteydessä opettajan käsitykseen oppilaan tavoitteellisuudesta (kognitiivisesta kyvykkyydestä, motivaatiosta ja kypsyydestä) ja opetettavuudesta, minkä on todettu olevan yhteydessä edelleen opettajan pedagogiseen päätöksentekoon eli tapaan, miten hän opettaa tiettyä oppilasta, miten hän kommunikoi oppilaan kanssa sekä siihen, miten paljon hän pitää kyseisestä oppilaasta. Näiden mekanismien kautta temperamentilla on todettu olevan yhteyttä myös oppilaan saamiin arvosanoihin.

Tutkimuksessa tarkasteltiin matematiikan ja äidinkielenopettajien käsityksiä oppilaan temperamentista, tavoitteellisuudesta ja opetettavuudesta sekä opettajan käsitysten yhteyttä opettajan oppilaalle antamiin matematiikan ja äidinkielen kouluarvosanoihin suomalaisessa väestöpohjaisessa aineistossa. Oppilaan ja opettajan sukupuolta ja opettajan ikää tarkasteltiin sekä kontrolloitua väliin tulevana muuttujina. Oppilaan omaa temperamenttiarviota tarkasteltiin kontrollimuuttujana. Tutkimusaineisto koostui peruskoulun yhdeksännen luokan oppilaista ($n = 3212$) ja heidän matematiikan ja äidinkielen opettajistaan ($n = 221$). Temperamentti ja opetettavuus mitattiin TABC-R ja DOTS-R -mittareilla. Tavoitteisuus mitattiin tähän tutkimukseen kehitetyllä mittarilla. Kouluarvosanat saatiin oppilaiden viimeisimmästä todistuksesta.

Tulokset osoittivat, että (1) opettajan käsitys oppilaan synnynnäisestä temperamentista, tavoitteellisuudesta ja opetettavuudesta on yhteydessä oppilaan saamaan matematiikan ja äidinkielen arvosanaan, myös silloin, kun oppilaan oma temperamenttiarvio on kontrolloitu; (2) opettajan oppilaasta tekemä arvio on yhteydessä oppilaan ja opettajan sukupuoleen ja opettajan ikään; (3) opettajan oppilaasta tekemä temperamenttiarvio ja oppilaan oma temperamenttiarvio näyttävät selittävän ja mittaavan eri asioita; ja (4) opettajan käsitykset oppilaan synnynnäisestä temperamentista, tavoitteellisuudesta ja opetettavuudesta muokkaavat opettajan antamaa arvosanaa. Koska kyseessä on poikkileikkaustutkimus, eivät tulokset anna perusteita päätelmille kausaliteetista.

Tulokset viittaavat siihen, että nykyisen koulujärjestelmän arviointikäytäntöä, jossa oppilaan oppimisprosessi ja tuntiaktiivisuus ovat keskeisessä asemassa, olisi suositeltavaa arvioida uuden temperamenttietoisen pedagogiikan valossa. Lisäksi, tulokset suosittavat temperamenttietoista pedagogiaa osaksi opettajankoulutusta, jotta opettajilla olisi välineitä (a) tunnistaa oppilaan synnynnäinen temperamentti; (b) erottaa se kognitiivisesta kyvykkyydestä, motivaatiosta ja kypsyydestä; ja (c) mahdollistaa oppilaalle kouluarviointi, joka on riippumaton oppilaan temperamentista ja persoonallisuudesta.

Avainsanat: temperamentti, koulumenestys, kouluarvosanat, opetettavuus, tavoitteellisuus

Acknowledgements

I owe my warmest and deepest gratitude to my supervisor, Professor Liisa Keltikangas-Järvinen, who gave me this remarkable privilege to work under her knowledgeable and encouraging supervision and to take part in her exceptionally talented research group, ‘Personality and Well-Being’. I have had an opportunity to grow into a researcher under her safe and wise wings. It is touching how she continually believed in my skills, rejoiced in my success and spurred me forward through difficult times. By some secret intuition, she has always known when to push me ahead towards new challenges and when to give me that time and space necessary to internalize things and events. It has been a true understanding and respecting of my own temperament. She has determinedly and comprehensively educated and trained me in all areas of researching—including that significant tacit knowledge of scientific research work which usually only accumulates with growing experience. Her door has always been open to me, and I wish every doctoral student could have a true mentor like her. I’m forever grateful to you, Liisa.

To my other supervisors, Docent Mirka Hintsanen and Professor Niklas Ravaja, I also want to express my deep and warm gratitude. They have always been there to help and support me whenever I needed it. They have been ready to share all their scientific knowledge, and my scientific thinking and writing skills have developed under their precise, competent and inspiring guidance. Thank you.

I wish to thank University Instructor Jari Lipsanen, Docent Markus Jokela and Docent emeritus Erkki Komulainen for their patient statistical advice. Through the different phases of my studies, they gave me much of their time and shared with me their excellent statistical know-how, for which I am very grateful.

I also want to thank all my other colleagues from the ‘Personality and Well-Being’ research group for their friendship and for helping and supporting me in so many ways. Saija Alatupa, Taina Hintsu, Päivi Merjonen, Laura Pulkki-Råback, Christian Hakulinen, Kim Josefsson, Ilmari Määttänen, Aino Pitkänen, Tom Rosenström, Nadezda Chumaeva, Maria Törnroos, Kaisa Kaseva, Elli Oksman, and Mauri Niiniaho—I could not have wished for, nor imagined, a better and more enjoyable research group than you.

With regards to the beginning of my research career, I must thank three emeritus Professors at the University of Helsinki. After completing my master thesis, Professor emeritus Pertti Kansanen proposed that I continue with doctoral studies. Professor emeritus Kari Uusikylä invited me to work as his research assistant and to participate in his interesting ‘Creativity and Gifted-

ness' research group. Professor emeritus Matti Meri further promoted my full-time work as a doctoral student at the University of Helsinki. Under their guidance and support, many years ago, all this began.

I am deeply indebted to Professor emeritus Juhani Hytönen and Professor Jari Lavonen from the Faculty of Behavioural Sciences. In turns as Head of the Department of Teacher Education, their brave, wise and broad-minded decisions enabled me to continue my PhD process. Without their unflinching support, I do not think I would be at this point now. Thank you.

I also warmly thank my colleagues Katriina Maaranen and Sanna Patrikainen from the Department of Teacher Education for their emotional support and for sharing the ups and downs of researching—and life—over the years.

Academy Professor Erno Lehtinen and Professor Päivi Atjonen as the official reviewers of the thesis are sincerely and warmly thanked for their critical and valuable comments that helped me to improve it. Amanuensis Kari Perenius and reviser John Gage from the Language Centre have been of invaluable assistance with the layout and language of the manuscript.

I am deeply grateful to my mother Kyllikki and my departed father Pentti for their solid and boundless love and support throughout my life. Their positive attitude and passion for conscientious work as private entrepreneurs have been a wonderful foundation to build on. You have always believed in me, thank you. I also warmly thank Matti, who has been like a true grandfather to my children for many years. He has always been so kind and helpful and has taken excellent care of my leisure-time home—Homeskorpi—my beloved safe house, where I have been able to rest my soul and recharge my batteries in times of life's happiness and sorrows. I also wish to thank my two sisters, Virpi and Mervi, for the valuable time and loving care they have given to my children during all these years.

Finally, I address my special thanks to my lovely children—son Ilari and daughter Elina—for showing me the things that matter most in life. You are my sunshine and greatest achievement. I am so happy about you two charming individuals with your unique temperaments. I therefore dedicate this doctoral thesis to you.

Homeskorpi, September 2012.

List of original publications

This doctoral dissertation is based on the following original publications¹, which are referred to in the text by their Roman numerals (*Studies I–IV*).

- I Mullola, S., Ravaja, N., Lipsanen, J., Hirstiö-Snellman, P., Alatupa, S., & Keltikangas-Järvinen, L. (2010). Teacher-perceived temperament and educational competence as predictors of school grades. *Learning and Individual Differences, 20*(3), 209–214. Copyright © 2010 Elsevier Inc.
- II Mullola, S., Ravaja, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., & Keltikangas-Järvinen, L. (2012). Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability. *British Journal of Educational Psychology, 82*(2), 185–206. Copyright © 2011 The British Psychological Society.
- III Mullola, S., Jokela, M., Ravaja, N., Lipsanen, J., Hintsanen, M., Alatupa, S., & Keltikangas-Järvinen, L. (2011). Associations of student temperament and educational competence with academic achievement: The role of teacher age and teacher and student gender. *Teaching and Teacher Education, 27*(5), 942–951. Copyright © 2011 Elsevier Ltd.
- IV Mullola, S., Hintsanen, M., Jokela, M., Lipsanen, J., Alatupa, S., Ravaja, N., & Keltikangas-Järvinen, L. (2012). Associations between teacher-rated versus self-rated student temperament and school achievement. *Scandinavian Journal of Educational Research, 1–27, iFirst Article*. Copyright © 2012 Scandinavian Journal of Educational Research.

Abbreviations: DOTS-R, Revised Dimensions of Temperament Survey; TABC-R, Temperament Assessment Battery for Children Revised; ML, Mother language; Math, Mathematics; EC, Educational competence; GPA, Grade Point Average; FTSA, Finnish Study of Temperament and School Achievement.

¹ The original articles are reprinted with the kind permission of the copyright holders. For permission to reprint or use in any form contact the copyright holders of the articles mentioned below.

Contents

1 Introduction	1
1.1 Temperament and teachability in the school context	1
1.1.1 Definition of Temperament	2
1.1.2 Definition of Educational competence	3
1.1.3 Definition of Teachability.....	3
1.1.4 Goodness or poorness of fit	7
1.2 Gender differences in temperament and teachability	7
1.3 Societal and cultural appraisals and expectations	8
1.4 Temperament and school achievement	9
1.4.1 Temperament and learning	9
1.4.2 Teachability and school achievement.....	10
1.4.3 Student and teacher gender, and teacher age.....	11
1.5 Teacher-rated versus self-rated temperaments and school achievement.....	13
1.6 Subject-related achievement and temperament.....	14
1.7 Assessment in the Finnish education school system.....	14
2 Aims of the study	19
3 Methods	25
3.1 Design of the study and selection of participants.....	25
3.1.1 Design of the Finnish Study of Temperament and School Achievement (FTSA)	25
3.1.2 Description of the sample, and sample collection in the FTSA study.....	25
3.1.3 Participants in the present study	27
3.1.4 Procedure of the present study.....	28
3.2 Measures.....	29
3.2.1 Teacher-perceived student temperament (Study I, II, III, and IV).....	29
3.2.2 Student self-rated temperament (Study IV).....	31
3.2.3 Teachability (Study IV)	31
3.2.4 Educational competence (Study I, II, and III)	32

3.2.5	School grades (Study I, III, and IV)	32
3.3	Statistical analyses	33
4	Results	37
4.1	Teachers' perceptions of student temperament, educational competence, and teachability	37
4.1.1	Main associations of teacher and student gender, and teacher age with teacher-perceived temperament, educational competence, and teachability (Study II)	37
4.1.2	Interactive associations of teacher and student gender, and teacher age with teacher-perceived temperament, educational competence, and teachability (Study II)	40
4.1.3	Summary of the results of Study II	41
4.2	Associations of teacher-perceived temperament, educational competence, and teachability with school achievement	42
4.2.1	Main associations of teacher-perceived temperament and educational competence with teacher gender and age in relation to Mathematics and Mother language grades (Study III)	42
4.2.2	Moderating and mediating associations of educational competence in relation to Mathematics and Mother language grades (Study I).....	43
4.2.3	Interactive associations of teacher-perceived temperament and educational competence with teacher and student gender, and teacher age in relation to Mathematics and Mother language grades (Study III).....	48
4.2.4	Gender differences in Mathematics and Mother language grades (Study III).....	50
4.2.5	Summary of the results of studies I and III	51
4.3	Teacher-rated versus self-rated student temperament, teachability, and school achievement.....	52
4.3.1	Factor structures of teacher-rated versus self-rated temperament (Study IV).....	52
4.3.2	Mutually adjusted associations of teacher- and self-rated temperament and teachability with school grades (Study IV).....	54
4.3.3	Summary of the results of Study IV	57

5 Discussion	59
5.1 Are there gender differences in teachers’ perceptions of student temperament, educational competence, and teachability?.....	59
5.1.1 Does teacher gender matter?.....	60
5.1.2 Does teacher age matter?	61
5.2 Are teacher-perceived temperament, educational competence, and teachability associated with student school achievement?	62
5.2.1 What is the role of educational competence?	64
5.2.2 Do teacher and student gender, and teacher age matter?.....	65
5.3. Which is the stronger factor in student school achievement—teacher-rated or self-rated student temperament and teachability?.....	66
5.4. Are there subject-related differences in the association between teacher- and self-rated temperament and teachability with student school achievement?	71
5.5 Methodological considerations.....	72
6 Conclusions and Practical Implications	75
6.1 Individual and societal consequences of the current findings	75
6.1.1 Boys’ exclusion	75
6.1.2 Girls’ later underachievement.....	76
6.1.3 Fairness of assessment.....	76
6.2 Knowledge of temperament in teacher training	77
6.3 Temperament-conscious education	79
6.4 Other remarks	80
References	81
Original Publications	91

1 Introduction

1.1 Temperament and teachability in the school context

School achievement has far-reaching consequences for the educational pathways and later life trajectories of youth. It determines a student's success or failure in being eligible for further studies as well as later career choices. School success requires favourable accomplishments with respect to at least two different and important school demands: (a) academic performance and achievement and (b) socially appropriate interpersonal behaviour (Keogh, 1986, 2003). Student temperament (i.e., an individual's unique, innate way of approaching and reacting to the environment) and teacher-perceived teachability (i.e., a teacher's perception of the attributes of an ideal model student) have been shown to have a broad and significant influence on both these demands (for reviews, see Guerin, Gottfried, Oliver, & Thomas, 2003; Keogh, 2003; Kristal, 2005; Strelau, 1998). Temperament has been shown to moderate learning in many situations by either facilitating or impeding certain learning strategies, learning processes and the successful completion of tasks (Davis & Carr, 2002; Guerin, Gottfried, Oliver, & Thomas, 1994; Martin, 1994). Student temperament and teachability have also been associated with student-teacher interactions (DiLalla, Marcus, & Wright-Phillips, 2004; Rudasill & Rimm-Kaufman, 2009), teacher expectations (Martin, 1989a, Martin, 1989b; Martin, Nagle, & Paget, 1983; Stipek, 2002) and educational decisions regarding a given student (Keogh, 1989; Pullis & Cadwell, 1982; Rothbart & Jones, 1998). Consequently, student temperament has been demonstrated to be an influential factor in predicting school success and academic outcomes measured by both standardized achievement tests and teacher-rated school grades (Guerin et al., 2003; Martin, 1989a; Martin & Holbrook, 1985; Martin, Olejnik, & Gaddis, 1994).

Teachers play a central role in being students' 'significant adults', and are therefore in a key position to influence students' school adjustment, achievements, and self-concepts (Guerin et al., 2003; Keogh, 2003; Kristal, 2005; Martin, 1992, 1994; Martin et al., 1994). Insofar as school achievement is intended to refer to the learning potential and academic performance of the student, it is important to minimize the 'external' influences that may bias the assessment of school achievement, e.g. whether students act or complete the learning processes in all respects as their teacher would like them to. Increasing teachers' knowledge of temperament and its consequences in the school context might help teachers to make more objective assessments of academic

performance. This would ensure equitable treatment and evaluation for all students, independent of their individual differences in characteristics such as temperament and personality.

1.1.1 Definition of Temperament

Temperament is defined as the “how of behaviour”, reflecting an individual’s unique responses to his/her environment (Thomas & Chess, 1977). It differs from ability (the “what and how well of behaviour”) and from motivation (the “why of behaviour”) (Thomas & Chess, 1977), being also relatively independent of cognitions (Oliver, Guerin, & Gottfried, 2007; Thomas & Chess, 1977). Only low to moderate correlations between measures of IQ and temperament have been found (for reviews, see Guerin et al., 2003; Kristal, 2005; Strelau, 1998), or IQ has been shown to moderate the relation between temperament and school achievement (Oliver et al., 2007). Hintsanen and colleagues (Hintsanen, Alatupa, Jokela, Lipsanen, Hintsanen, & Leino, 2012) recently demonstrated that temperament was associated with school achievement even when student motivation and standardized cognitive ability test performance were taken into account.

Temperament contributes to the uniqueness of individuals. Although there are currently several competing theories and definitions of temperament (see Goldsmith et al., 1987; Strelau, 1998), a consensus exists that temperament refers to a biologically based, innate behavioural style, which become evident in early childhood and is relatively stable across different situations and over the course of time (Bates, 1989; Buss & Plomin, 1975, 1984; Goldsmith, Buss, Plomin, Rothbart, Thomas, Chess et al., 1987; Goldsmith, Lemery, Aksan, & Buss, 2000; Thomas & Chess, 1977; Rothbart, 1989). Temperament is seen as raw material that forms an emotional basis for the later development of personality (Angleitner & Ostendorf, 1994; Goldsmith et al., 2000), which in turn includes cognitions about self and others as well as social factors like values, attitudes, and coping strategies (for reviews, see Posner & Rothbart, 2007; Rothbart, 2007; Rothbart & Bates, 2006; Strelau, 1998).

In everyday school life, teachers mostly recognize temperament through a “we know it when we see it” attitude (Keogh, 2003). However, research has shown that this practical and subjective perception by a teacher may lead to a ‘halo effect’ between temperament, EC, and school grades; that is, because of temperament, a teacher perceives some students as more mature than others. In addition, on the basis of a student’s working style, e.g. because she/he works hard and is persistent and adaptable, a teacher may subconsciously be inclined to give that student a higher grade than what the performance actu-

ally warrants. However, a student's working style refers to his/her innate temperament, not to cognitive ability. The same teacher may thus not adequately acknowledge a student perceived as less adaptable and less persistent and more susceptible to being interrupted (Keogh, 1982; Martin et al., 1994). Students have also been found to perform significantly worse on subject tests when the teacher of that subject has viewed them negatively (i.e., as disruptive, inattentive, or rarely completing homework) (Dee, 2005).

Given that some temperament traits facilitate certain learning processes and successful completion of tasks, some students are perceived as more "teachable" than others. Contrarily, other temperament traits are more impeding and harmful in relation to a student's learning strategies. This "appropriateness" is linked to the school context due to the nature of the school curriculum and classroom setting. The relevance of the same temperament traits may, however, completely change in other contexts and environments outside school.

1.1.2 Definition of Educational competence

In the current study, student *educational competence* (EC) refers to the three significant indicators of students' school performance and school success: cognitive ability, motivation and maturity (readiness for school attendance) as perceived by a teacher. Previous studies have adopted slightly different ways to measure the concept of EC. Keogh (Keogh, 1983) has used the concept in terms of teachers' ratings of pupils' ability, motivation, and social and academic competence as well as students' management and monitoring needs in the classroom. On the contrary, Shin (Shin, 2003) used the level of student reading skill as an indicator to measure students' EC. Although different studies have used diverse measures as indicators of student EC, in all of them EC refers to a student's general school readiness and capacity to receive teaching, absorb information and adjust to working in the school environment. In the current study, a student's EC reflects the teacher's perception of student EC and its three components, not objectively measured student's cognitive ability.

1.1.3 Definition of Teachability

Student *teachability* (Keogh, 1982, 1983, 1994, 2003), illustrated in Figure 1, reflects teachers' perceptions of the attributes of an ideal model student, one who has a compliant ability to receive teaching by reacting to it in the proper way the teacher desires. Teachability is composed of three factors: task orientation, personal-social flexibility, and reactivity (Keogh, 1982; Keogh, Pullis,

& Cadwell, 1982). These three factors are temperamental characteristics that are consistent with the definition of temperament². A student's teachability refers to a positive 'school temperament' (Martin, 1989a). The relative importance of each of the three factors may vary between school grade levels and different teachers, but generally there is strong agreement about the attributes of the model and of ideal pupils among teachers' perceptions, which are not absolutely based on students' actual ability (Keogh, 1982, 1986).

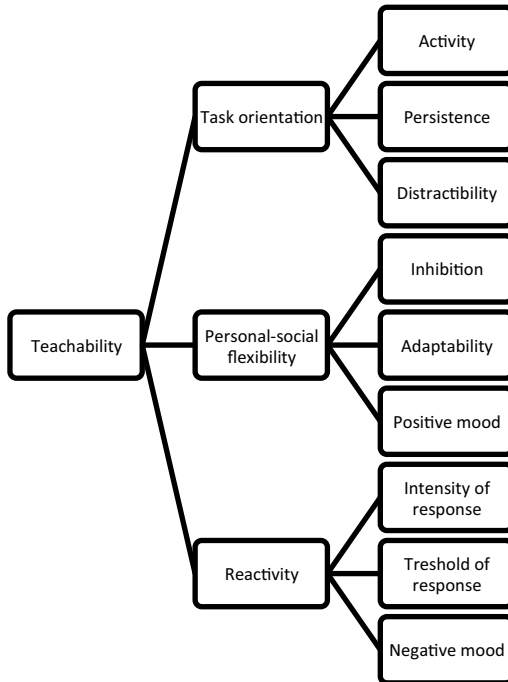


Figure 1. The concept of 'Teachability' according to Keogh's theory (Keogh, 1982, 1983, 1994, 2003)

Despite significant correlations between temperament and teachability (Keogh, 1982, 1986, 1994, 2003), these two conceptions are not synonymous. Teachability is viewed as a more comprehensive construct compared with temperament (Keogh, 1982, 1986, 1994, 2003), whereas temperamental

² Based on previous literature, individuals' temperamental differences are described by such concepts as *characteristic, dimension and trait*, which are used interchangeably also in this study.

variation is a major facilitator within teachability. The concept of teachability was initially a psychological rather than pedagogical one, and more related to developmental psychology than to education. It is consistent with Thomas & Chess's (Thomas & Chess, 1977) concept of temperamentally "easy" versus temperamentally "difficult" children (Keogh, 1982).

The first factor of teachability is *task orientation*, which is composed of three temperamental dimensions: activity, persistence and distractibility (Keogh, 1983). Activity refers to the frequency and intensity of motor activity (gross motor vigour and tempo), while persistence refers to attention span and the tendency to continue seeking a solution to difficult learning or performance problems (Martin, 1989a; Windle & Lerner, 1986). The third component of task orientation, distractibility, means the ease with which a student's attention can be interrupted by low-level environmental stimuli (Martin, 1989a; Windle & Lerner, 1986). These traits have been related to working styles, such as attention focusing, the vigour and eagerness with which the student approaches a learning task, and the learning situations and assignments he or she engages in (Caspi, 1998; Caspi & Shiner, 2006; Martin, 1989a; Kristal, 2005). Consequently, task orientation has been found to correlate significantly positively with the "school appropriate behaviour" component (i.e., follows directions, is attentive and alert to classroom events, and finishes tasks on time) (Keogh, 1994).

Students with high task orientation (i.e., low activity, high persistence, and low distractibility) have a better attention span and tolerance to frustration when learning new skills as well as an ability to complete academic activities (e.g., homework) at a higher rate than their peers (Martin & Bridger, 1999). Teachers perceive these students very positively in academic settings, and as more mature and capable beyond their actual age and performance (Martin & Bridger, 1999). On the other hand, students with low task orientation (i.e., high activity, low persistence, and high distractibility) are easily "sidetracked" from the given learning tasks, particularly difficult ones. They have problems maintaining attention and sitting still during long periods, which require high concentration and involve occasions of frustration (Martin & Bridger, 1999). These students are perceived by their teachers as less mature, as well as disruptive and histrionic.

The second factor of teachability is *personal-social flexibility*, which is comprised of three temperamental dimensions: approach-withdrawal/inhibition, adaptability, and level of positive mood (Keogh, 1994, 2003). It contributes to changes and transitions during the school day. The approach-withdrawal/inhibition dimension refers to a student's initial reaction of either approaching or withdrawing from new things and social situations, such as

people, situations and places. Adaptability reflects how easily and quickly a student adjusts to altered circumstances and transitions, and whether he/she needs time to prepare or even plan a schedule in order to move from one situation to another. Mood is a basic quality of mental disposition and may vary between more positive (glad or cheerful) and more negative (grumpy or somber).

Students with high personal-social flexibility (i.e., high tendency to approach new things, high adaptability, and positive mood) are characterized as happy, flexible, and direct, who make transitions easily, like new things, and do not need a schedule (Keogh, 1994; Kristal, 2005). Conversely, students with low personal-social flexibility (i.e., high withdrawal/inhibition, low adaptability, and negative mood) are described as slow-adapting and “clingy” students, whose initial response to new things is to withdraw. Parents and teachers generally describe them as shy children, who are most comfortable with familiar things and situations (Keogh, 2003; Kristal, 2005).

The third factor of teachability is *reactivity*, which consists of three temperamental dimensions: threshold of response (i.e., the intensity and strength of the stimulation that is needed to arouse a reaction), level of negative mood, and intensity of response (i.e., the expressive and reactive energy of a reaction, whether happy, sad, or angry) (Keogh, 1994, 2003; Kristal, 2005). The reactivity factor contributes to teachers’ opinions of students, such as how much a teacher likes a student (Keogh, 1994), and is also associated with a teacher’s perception of a student’s personal-social competencies (Keogh, 2003) and social position in the class (Hintsanen, Alatupa, Pullmann, Hirstiö-Snellman, & Keltikangas-Järvinen, 2010).

Teachers view highly reactive students (i.e., over-reactive with a low threshold of response, high negative emotionality, and high intensity of response) as less able (Lerner, Lerner, & Zabski, 1985), irritable and “prickly”, and difficult to teach (Keogh, 2003). Students low in reactivity (i.e., high threshold of response, positive mood, and low intensity of response), and with a combination of positive mood and moderate intensity, are viewed by teachers as more teachable pupils (Keogh, 1982, 1994). This combination of positive mood and moderate intensity has also been found to be associated with personal-social flexibility, particularly when it has not merely been a question of the extremes of intensity, but a combination of intensity and quality of mood (Keogh, 1982, 1994).

To summarize, the following temperament dimensions are of great importance in school settings according to Keogh’s theory of teachability (Keogh, 1982, 1983, 1994, 2003): activity, persistence, distractibility, inhibition,

adaptability, quality of mood, threshold of response, negative emotionality, and intensity of response.

1.1.4 Goodness or poorness of fit

Goodness or poorness of fit is the fit between temperament and school context, meaning the compatibility (i.e., goodness of fit) or dissonance (i.e., poorness of fit) between learning circumstances offered by the school environment and the student's own behavioural characteristics, working styles and capacities (Chess & Thomas, 1999; Thomas & Chess, 1977). It is the central concept in the temperament theory and approach developed and introduced by Thomas & Chess (Thomas & Chess, 1977). In terms of the temperament approach, temperament traits are neither good nor bad, but changes in the social environment may cause changes in the expression of emotional reactions aroused by temperament (Buss & Plomin, 1975; Chess & Thomas, 1999; Rothbart & Jones, 1998; Strelau, 1998; Thomas & Chess, 1977). "Goodness of fit results when the properties of the environment and its expectations and demands are in accord with the organism's own capacities, characteristics, and style of behaving" (Chess & Thomas, 1999, p.3). A continual imbalance between school expectations and a student's innate temperament may, according to the goodness of fit concept, result in enduring stress (Chess & Thomas, 1999; Thomas & Chess, 1977, 1980). Therefore a key factor in understanding the role of temperament in school achievement is the fit between temperament and school context.

'Goodness of fit' is seen as an interactive approach between students and the classroom environment, where both students' and teachers' characteristics and classroom circumstances must be taken into account (Pullis, 1989). There are at least three important aspects that interact with students' temperaments and may lead to either a good or poor fit in classroom settings (Keogh, 2003, pp. 31–32): (1) the content and nature of the curriculum; (2) the organization and management of space, time, and resources; and (3) the nature of the interactions between students, peers, and teachers. Overall, the goodness of fit should be made equally possible and achievable for all types of temperaments and both genders, assuming that all students should have equal opportunities to succeed according to their actual ability.

1.2 Gender differences in temperament and teachability

Concerning the above-mentioned school-related temperament traits, researchers have found a number of gender differences. Boys have been found to be less persistent and flexible (Sanson, Smart, Prior, Oberklaid, & Pedlow,

1994), and more active (Ahadi, Rothbart, & Ye, 1993; Coplan, Barber, & Lagacé-Séguin, 1999; Eaton & Enns, 1986; Maccoby & Jacklin, 1974) and distractible (Mendez, McDermott, & Fantuzzo, 2002) than girls, who in turn have been found to be less hyperactive (Kwok, Hughes, & Luo, 2007), and, as an indicator of positive affect (Rothbart, 1981), to smile more (Ahadi et al., 1993; Hall & Halberstadt, 1986; LaFrance, Hecht, & Paluck, 2003; Zhou, Lengua, & Wang, 2009). Regarding the greater male activeness, the difference appears after the first year of life and increases with age (Eaton & Enns, 1986; Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006; Maccoby & Jacklin, 1974).

Teachers have rated girls higher in inhibitory control and attention focusing and lower in irritability than boys (Zhou et al., 2009). Large meta-analytically found gender difference in effortful control (i.e., attention span, inhibitory control, and perceptual sensitivity) favouring girls (Else-Quest et al., 2006) has been concluded to be related to greater male incidence of attention deficit and/or hyperactivity problems (Nigg, Goldsmith, & Sachek, 2004). Furthermore, the same meta-analytical study showed moderate gender differences in surgency (i.e., high activity, high impulsivity, and high-intensity pleasure), with boys having higher surgency, as well as negligible gender differences in negative affectivity (Else-Quest et al., 2006). The results suggest that girls overall may have a better ability to control or allocate their attention, control inappropriate behaviours, and perceive low-intensity environmental stimuli that may lead to a better awareness of subtle environmental changes (Else-Quest et al., 2006).

Boys' teachability is seen to be lower than girls' (Keogh, 1994; Van Houtte, 2007). Teachers tend to perceive boys as having lower levels of task orientation (Keogh, 1994) and general attention focusing (Else-Quest et al., 2006; Zhou et al., 2009). Furthermore, boys are perceived as being higher in inhibition (Keogh, 1994) and reactivity (Else-Quest et al., 2006) and lower in positive mood compared to girls (Keogh, 1994). Boys are also rated as less flexible and adaptable (Guerin & Gottfried, 1994; Keogh, 1994; Sanson et al., 1994) and as having difficulties in showing appropriate school behaviour in adjusting to classroom demands (Keogh, 1989, 1994).

1.3 Societal and cultural appraisals and expectations

The cultural and socio-economic context in which children develop shapes the manifestation and expression of temperamental characteristics (Ahadi et al., 1993; Kohnstamm, 1989; Lewis, 1989; Rothbart & Bates, 2006; Windle, Iwawaki, & Lerner, 1988; Yang, 1986) and the way different temperament traits are perceived and valued (Kerr, 2001; Zhou et al., 2009). The extent to

which a culture values or accepts certain behaviours may drive the reinforcement or punishment of these behaviours, resulting in different developmental outcomes (e.g., in terms of social skills, peer relationships, and later adjustment) (Ahadi et al., 1993; Boivin, Hymel, & Bukowski, 1995; Chen, Rubin, & Li, 1995; Chen, Rubin, & Sun, 1992; Zhou et al., 2009), and long-term consequences concerning temperament-related inter-personal processes (e.g., the age of marrying and having children) (Caspi, Elder, & Bem, 1988; Else-Quest et al., 2006).

Cultural norms and expectations have been shown to be gender-related; i.e., a similar temperament may be viewed as more or less appropriate or desirable in boys and girls (Ahadi et al., 1993; Kerr, 2001; Kerr, Lambert, & Bem, 1996; Kohnstamm, 1989; Radke-Yarrow, Richters, & Wilson, 1988; Stevenson-Hinde, 1988; Stevenson-Hinde & Hinde, 1986). This may in turn, and in all likelihood, moderate teachers' perceptions of an ideal and "teachable" student (Keogh, 2003; Kerr, 2001; Kerr et al., 1996). These processes become activated and visible especially in the school environment, where the manners and customs of behaviour are culturally bounded and slow to change.

Teacher ratings have been concluded to be even more gender typed than parents' ratings because teachers frequently see students interacting in peer groups with the same gender, which has been seen as magnifying gender role differences (Else-Quest et al., 2006; Maccoby, 1990). Teachers of an opposite gender have been found to be more susceptible to perceiving a student as inattentive, disruptive, or prone to omit homework than teachers of the same gender as the student (Dee, 2005). On the other hand, teachers may view girls' quiet and persistent ways of working as more mature and teachable than those of energetic and active boys. They may also view the behaviour of inflexible boys as difficultness, because boys may not as easily adjust to the many changes in classroom demands (Keogh, 1989, 2003). There is, however, no research-based evidence of gender differences in adaptability (Else-Quest et al., 2006), defined by Thomas and Chess (Thomas & Chess, 1977) as referring to a student's ability to adjust to repeated changes in school.

1.4 Temperament and school achievement

1.4.1 Temperament and learning

Temperament may affect a student's school achievement through his/her working style (Guerin et al., 1994; Kristal, 2005; Rothbart & Jones, 1998) and selected working strategies (Davis & Carr, 2002). Temperament is also related to a student's interests and general enjoyment of school (Elliot &

Trash, 2002; Guerin et al., 2003), as well as energy and willingness to approach certain learning tasks (Caspi, 1998; Caspi & Shiner, 2006; Keogh, 2003; Kristal, 2005; Martin, 1989a). Temperamental activity level, persistence, distractibility and flexibility in particular may mediate a student's successful fulfillment of school tasks by either assisting or complicating the student's learning process (Guerin et al., 1994, 2003; Kristal, 2005; Orth & Martin, 1994). Furthermore, these associations may be partly gender-related (Davis & Carr, 2002; Ham et al., 2006).

Boys' low temperamental impulsivity and girls' low temperamental inhibition have been found to be associated with successful retrieval strategies in problem-solving (Davis & Carr, 2002). Boys with higher impulsivity tend to use a more manipulative approach in problem-solving tasks than boys with lower impulsivity (Davis & Carr, 2002), whereas among girls no associations between impulsivity and strategy use have been found. Boys with higher self-directedness have achieved higher GPAs (grade point averages) than boys with lower self-directedness (Ham et al., 2006). Conversely, girls with higher harm avoidance have achieved higher GPAs than girls with lower harm avoidance (Ham et al., 2006). In the same study, students with higher GPAs (whether girls or boys) were differentiated as higher in persistence and lower in novelty seeking compared with students with lower GPAs (Ham et al., 2006).

Temperament is defined to be rather independent of student motivation and maturity (Thomas & Chess, 1977) and has been found to only modestly correlate with intelligence (for reviews, see Guerin et al., 2003; Keogh, 2003, Kristal, 2005; Strelau, 1998) and other cognitive functions (Oliver et al., 2007; Thomas & Chess, 1977). For example, the results of Davis and Carr (Davis & Carr, 2002) revealed no evidence of an association between temperament and cognitive strategy use in Math problem-solving tasks, and this was accurate for both genders. Consequently, temperament is not likely to be associated with student cognitive abilities, and thus the association of temperament with school achievement might be seen as bias.

1.4.2 Teachability and school achievement

Students with low task orientation, low personal-social flexibility and high reactivity are consistently perceived by their teachers as less teachable and as having poor EC (Keogh, 1983, 2003). Compared with the results of standardized cognitive tests, teachers underestimate the intelligence (Keogh, 1982; Martin & Holbrook, 1985), motivation, maturity and cognitive ability (Keogh, 1994; Lerner et al., 1985; Pullis & Cadwell, 1982) of students with low task orientation and perceive them as lazy students who shirk their responsi-

bility and are difficult to teach (Guerin et al., 2003). Low task orientation in particular correlates with lower grade point averages (GPAs) (Ham et al., 2006) and lower teacher-rated school grades (Martin, 1989a; Martin et al., 1994).

Studies have also shown that students with high personal-social flexibility (positive mood and high adaptability) receiving higher grades than might be expected on the basis of standardized achievement tests (Keogh, 1982, 1994; Lerner et al., 1985; Martin & Holbrook, 1985). Further, students with better teachability with respect to teacher demands have received higher teacher ratings for academic ability and adjustment as well as higher GPAs than students with poorer teachability (Lerner et al., 1985). On the other hand, students' high reactivity has been found to be associated with teachers' estimates of pupils' lower school adjustment, higher management needs and lower performance on a school readiness test (Keogh, 1983, 1994, 2003). However, already moderate reactivity allows students to be viewed as more teachable (Keogh, 1982, 1994). Overall, students' reactivity has been found to be associated with teachers' views of students' personal-social competencies (Keogh, 1994, 2003).

Teachers have also been found to show less trust in students they perceive as less teachable (Van Houtte, 2007). Consequently, children with poor student-teacher relationships, marked by conflict and dependency, have received lower grades at school (e.g., DiLalla et al., 2004; Hamre & Pianta, 2001).

1.4.3 Student and teacher gender, and teacher age

In the western world a lively conversation is taking place about the number of male teachers in teacher training and as practitioners in education systems (Carrington & Skelton, 2003; Cushman, 2008; Drudy, 2008; Francis, 2008). It has been suggested that the feminization of the teaching profession explains boys' underachievement (Carrington & McPhee, 2008; Skelton, 2002) and that boys would need male teachers to do better (Dee, 2007; Drudy, 2008) or to have a positive male role model (Cushman, 2008; Francis, 2008). This conversation is a central issue in Finland (Lahelma, 2000) but topical also in other OECD countries where female teachers comprise the largest proportion of secondary school teachers (Carrington & Skelton, 2003; Cushman, 2008; Drudy, 2008; Francis, 2008).

Teachers' female gender has previously been suggested to have a positive impact on students' reading performance (Lam, Tse, Lam, & Loh, 2010), language learning (Chudgar & Sankar, 2008) and both genders' general achievement (Krieg, 2005; UNESCO, 2005). There are also dissimilar and conflicting findings of a positive male teacher effect on students' Math learn-

ing (Warwick & Jatoi, 1994), same-gender effect on students' achievement (Michaelowa, 2001) and no gender effect on either Math (Chudgar & Sankar, 2008) or any other subject outcomes (Driessen, 2007).

Bettinger and Long (Bettinger & Long, 2005) and Dee (Dee, 2007) found that same gender teachers increased student's interest and engagement in a teacher's subject. Female teachers' influence was strong in several subjects, particularly Math and statistics but among males only in education (Bettinger & Long, 2005), revealing a positive role-model effect in such fields where that gender is underrepresented.

There is also evidence of same-gender matching having no effect on student's outcomes (Driessen, 2007; Ehrenberg, Goldhaber, & Brewer, 1995; Lahelma, 2000) and also of boys' school attitudes being even more positive when taught by female teachers (Carrington, Tymms, & Merrell, 2008; Sokal, Katz, Chaszewski, & Wojcik, 2007). On the other hand, based on teachers' practical perceptions and partly tacit assumptions, male teachers have been assumed to practice somewhat more relaxed pedagogy with their students compared to female teachers, or to teach in ways that are more appealing to boys (Ashley & Lee, 2003; Francis, 2008; Skelton, 2002).

Two extensive international studies have investigated the effect of teachers' gender in the context of teaching and teacher training. Their results showed female teachers perceiving their students as less teachable and having less trust in them (Van Houtte, 2007), as well as punishing actions with recidivism more than male teachers (Salvano-Pardieu, Fontaine, Bouazzaoui, & Florer, 2009). Male teachers, in turn, preferred girl students in terms of both teachability and trust (Van Houtte, 2007) and punished misbehaviour by academically good students more than female teachers (Salvano-Pardieu et al., 2009). Further, older teachers showed more trust in their students compared with younger teachers independent of teacher-perceived teachability (Van Houtte, 2007). Teachers' age has also found to be associated with the motivational factors of teacher professionalization (Hildebrandt & Eom, 2011), teachers' emotional responses to educational change (Hargeaves, 2005), and teachers' severity of punishment (Salvano-Pardieu et al., 2009). However, no previous study has taken student temperament into account in any of these contexts. In addition to scant studies on the effect of teacher gender and age on the associations between students' temperaments and outcomes, there is a lack of knowledge on how teacher and student of the same or different gender affects perceptions and outcomes. That is, does it matter whether the teacher and the student are of the same or different gender?

1.5 Teacher-rated versus self-rated temperaments and school achievement

The comparison of the relevance of the method used here, i.e., teacher-rated versus self-rated temperaments in relation to school achievement, has received little attention in the literature. Most studies have been based on teacher-, parent-, or self-ratings examined separately and in different samples (for reviews, see Guerin et al., 2003; Keogh, 2003; Kristal, 2005; Strelau, 1998; Zhou, Main, & Wang, 2010). The discrepancies and relative strengths of the different methods and effects of teacher- and self-rated temperament in predicting school achievement in the same sample have therefore remained virtually unexplored. Hintsanen and colleagues (Hintsanen et al., 2012) recently demonstrated the significance of the method used here relative to observed results. Temperamental activity and negative emotionality were significant factors of Math grade only when rated by a teacher, whereas inhibition was a significant factor only when self-rated.

The results from temperament ratings by teachers, parents and students may not be in agreement with each other because each represents a different viewpoint and may be influenced by different types of biases (for a review, see Achenbach, Dumenci, & Rescorla, 2002; Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Achenbach, McConaughy, & Howell, 1987; Keogh, 2003; Vazire & Mehl, 2008). In fact, it is not merely a question of discrepancies between the used methods, but also the different methods might measure somewhat different phenomena. Self-rated temperament might be related more to student's self-esteem (Klein, 1995) or as previously suggested, it might be more affected by views held by their peers (Hintsanen et al., 2012). Given that the teacher interacts with and observes the students in the school environment in certain repeated situations, the teacher ratings may be strongly affected by a teacher's opinions and attitudes and less by the distinct traits of the student (Field & Greenberg, 1982; Stipek, 2002). In particular, temperamental task orientation contributes to how teachers perceive student performance and achievement in various tasks (Keogh, 1982; Martin et al., 1994; Stipek, 2002). This may lead to the development of a 'halo effect', where the teacher unconsciously mixes his/her perceptions of student temperament with cognitive abilities and school performance and transfers this perception to the assigning of school grades (Keogh, 2003).

In addition, the correlations between temperament and cognitive abilities have been stronger for teacher ratings as compared to parent ratings (Keogh, 2003) and student self-ratings (Hintsanen et al., 2012), although teacher and parent ratings in particular may be biased by the selective expectations of teachers concerning student achievement (Keogh, 2003; Martin & Holbrook,

1985; Martin et al., 1994; Kristal, 2005). The association between teacher-reported temperament and achievement expectations is particularly apparent if the student temperament profile is viewed as being negative instead of positive or neutral (Keogh, 1982, 2003). Such perceptions may be less structured and specific than people's perceptions of personality in general. However, according to previous studies, it is difficult to estimate the degree to which correlations between school achievement and teacher-rated temperament traits reflect a student's behaviour versus a teacher's perceptions (Jussim & Harber, 2005).

1.6 Subject-related achievement and temperament

The role of temperament in school achievement may vary according to school subjects and the demands set by the different subjects. Different school subjects favour different study techniques. Some, such as Math, may call for greater engagement and ability to concentrate (related to high temperamental task orientation) than others (Keogh, 1982, 1983, 1986). Other subject-related tasks may necessitate more articulacy, creativity with new problem-solving strategies and/or engagement in new situations (related to temperamental flexibility and reactivity) (Keogh, 1982, 1983, 1986). In Mother language (ML) instruction, for example, the curriculum involves various working methods in versatile domains, such as grammar, literature and reading, creative writing, and articulacy and drama (FNBE, 2004). These domains contain both fixed as well as free forms of study content, working styles and methods. The Math curriculum, on the other hand, is more divided and focused firmly on its mathematical approach and content, and the working methods may be more limited and convergent.

Previous studies have found temperament to be more strongly related to ML than to Math (Guerin et al., 1994; Martin, 1989a, 1989b; Martin & Holbrook, 1985; Newman, Noel, Chen, & Matsopoulos, 1998; Strelau, 1998). However, opposite findings also exist (Maziade, Cote, Boutin, Boudreault, & Thivierge, 1986) as well as findings where teacher-rated temperament has been related to both subjects (Rudasill, Gallagher, & White, 2010; Strelau, 1998). Due to these inconsistent findings, further research is needed to investigate temperament's relationship with Math and ML.

1.7 Assessment in the Finnish education school system

The Finnish education system, shown in Figure 2 (FNBE, 2004), consists of nine years of compulsory schooling between the age of 7 and 15 (six years at primary school and three years at lower secondary school). The whole age

group can be easily contacted and sampled because there are no private schools or parallel school systems. Teachers are similarly educated and all schools follow the same national curriculum, Finnish comprehensive school is a most appropriate “real-life laboratory”, and for the following reasons. Approximately 97% of students in each age group go through public comprehensive school in regular classes (of which approximately 7% are under special, individual supervision, 2% are in special, “tailored” classes, and less than 1% leave without completing their education). Following comprehensive school, almost all students continue on to either Senior High School (approximately 64%) or a Vocational Institution (approximately 30%), and less than 5% drop out (Tilastokeskus, 2011).

The Finnish National Board of Education (FNBE) has specified the national guidelines and principles for pupils’ assessment in the National Core Curriculum for Basic Education intended for pupils in compulsory education (FNBE, 2004). According to the FNBE, three main tasks are involved in students’ assessment: (1) “to guide and encourage studying and to help pupils in their learning process”, (2) to perform a “final assessment of basic education, on the basis of which pupils will be selected for further studies, when they leave comprehensive school”, and (3) “to develop the pupil’s capability for self-assessment” (FNBE, 2004/Pupil Assessment, p.1). Teacher’s continuous and positive feedback and students’ equal treatment are seen as important elements of the evaluation process. For all subjects and classes, teachers’ assessment work has been guided by descriptions of good performance (i.e., grade 8 or “good”) and criteria for final assessment at the end of ninth grade (FNBE, 2004). The criteria of “good performance” have been set for teachers as a tool and a support for assessment work. With the assistance of the criteria, the school grades and GPAs are based on a) examinations designed by an individual teacher, b) model examinations offered by the authors of the textbooks and workbooks, c) teachers’ perceptions of students’ participative activity, carefulness and conscientiousness with respect to assignments during school lessons and working periods (including homework), and d) teachers’ perceptions of student’s abilities to complete the required teaching-learning process in a teacher’s planning and expecting way.

The Finnish National Board of Education (FNBE, 2004) has advised teachers to take students’ lesson activity and learning process into account in school assessment; these may influence a student’s final grade by one grade, if the teacher considers it necessary. The national standardized tests are mainly suggestive and to be used only once or twice in a school year, and only for certain subjects, mostly Math, ML, and foreign languages. Further, the subject-related school grade generally contains a teacher’s perceptions of

a student’s enthusiasm and motivation. In practical terms this means that a student’s interest in a certain subject and his/her hobbies may raise the grade point even though the student’s cognitive knowledge or other skills might not correspondence with the required target level.

FORMAL EDUCATION IN FINLAND

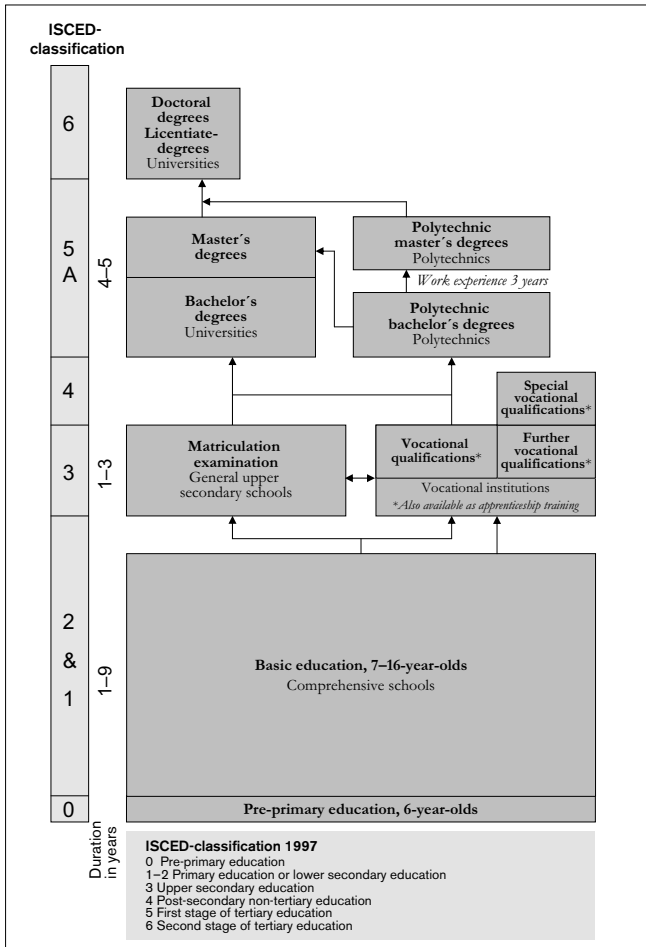


Figure 2. The construction of the formal education system in Finland.

Reference: Ministry of Education and Culture: Formal Education in Finland:
http://www.minedu.fi/export/sites/default/OPM/Koulutus/koulutusjaerjestelmae/liitteet/finnish_education.pdf

The final grade point is a combination of these above-mentioned elements assigned by an individual teacher. The element that a teacher primarily focuses on during assessments may vary between schools, classes and teachers. Among Finnish teachers, a student's progress in a particular learning process and the way the process is completed is important in assessment. Especially in elementary school, grades are supplemented with verbal assessments considering students' teacher-perceived abilities, working skills, adjustment to the required learning process and general adjustment to the school environment.

2 Aims of the study

The main aim of the study was to examine how the teacher perceives student innate temperament in the school context and whether teachers' perceptions of student temperament are associated with their perceptions of student educational competence (EC) (i.e., cognitive ability, motivation and maturity) and teachability. In addition, the aim was to examine whether teacher-perceived temperament, EC and teachability are related to student school achievement in terms of teacher-assigned school grades in Mathematics (Math) and Mother language (ML). Teacher and student gender and teacher age were examined as moderating variables, and student self-rated temperament as a control variable, in the association between teacher-perceived temperament and school achievement. Gender controls were implemented because previous studies have indicated that teachers' perceptions of student temperament might be gender-related due to societal and cultural appraisals and expectations (Ahadi et al., 1993; Kerr, 2001; Kerr, Lambert, & Bem, 1996; Kohnstamm, 1989; Radke-Yarrow, Richters, & Wilson, 1988; Stevenson-Hinde, 1988; Stevenson-Hinde & Hinde, 1986). This may moderate teachers' perceptions of an ideal and 'teachable' student in the school environment (Keogh, 2003; Kerr, 2001; Kerr et al., 1996), and have a further influence on teachers' assessment practices. In addition, previous research has indicated international concern over the feminization of the teaching profession, particularly considering boys' school performance and school well-being (Carrington & McPhee, 2008; Dee, 2007; Drudy, 2008; Lahelma, 2000; Skelton, 2002).

Four separate sub-studies were conducted. The focus and the research design of the current study are represented in Figure 3. The process and steps of the research are presented in Figure 4. Each study answered the question set for it, but raised some new research problems, which functioned as a building block and starting point for the next study.

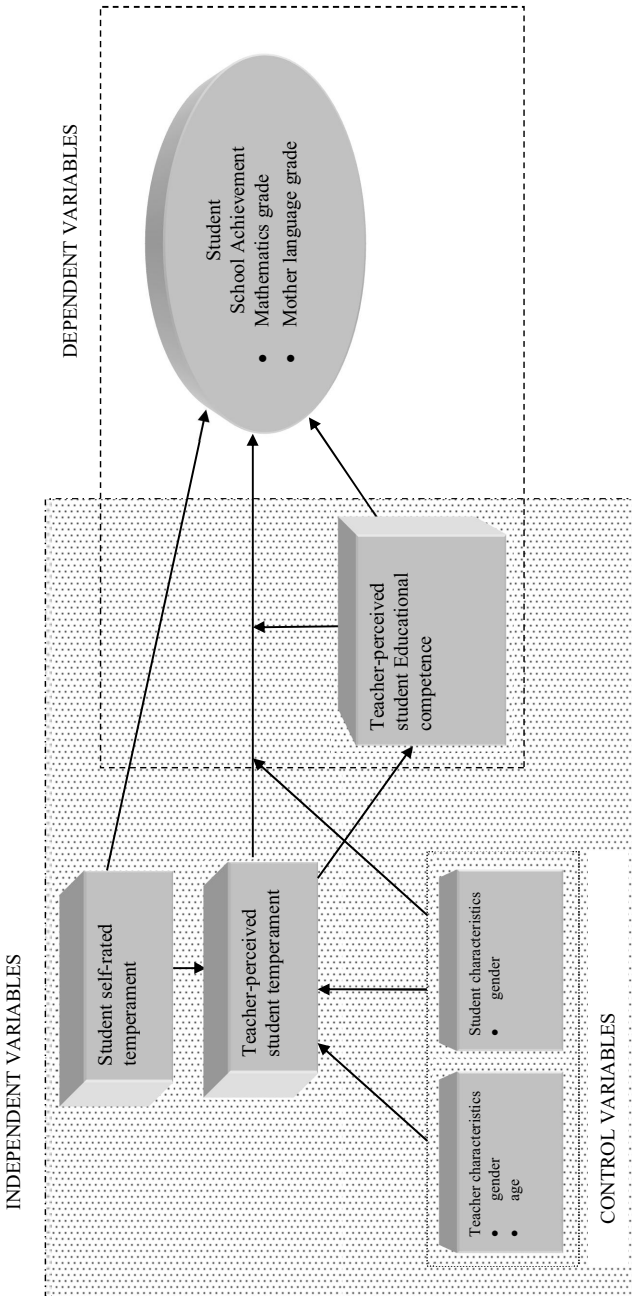


Figure 3. Focus of the current study.

Note. Although the associations are illustrated with arrows, the directions of the arrows are hypothetical, because cross-sectional research design does not allow the complete establishment of temporal directions and causality of the associations.

Study I

The aim of study I was to examine the associations of teacher-perceived temperament and EC (i.e., cognitive ability, motivation and maturity) with Math and ML grades, in a population-based sample of Finnish adolescents in their last year of comprehensive school. It was hypothesized that (a) temperament is directly related to teacher-perceived EC and school grades, (b) EC contributes as a mediator and/or moderator of the association between temperament and school grades, and (c) because Math and ML require different skills, students would be differently perceived by their respective teachers. Study I functioned as an explorative background study, which formed the preconditions for the further studies and research questions developed phase by phase.

Study II

Study II examined gender differences in teachers' perceptions of student's temperament, EC, and teachability, and whether there is a significant and systematic same-gender or different-gender association between teachers and students in this relationship. In addition the role of teacher age in this association was investigated. It was expected that (1) in general, teachers would perceive boys as lower in task orientation, EC, and teachability than girls. In addition, it was assumed that (2) male teachers would view both genders more positively and more highly in task orientation, EC, and teachability than female teachers, (3) female teachers would view girls more highly in task orientation, EC, and teachability than boys, and (4) in general, teachers' ageing would increase their negative views of students' temperament, especially concerning task orientation, and decrease their negative views of students' EC and teachability.

Study III

The purpose of study III was to examine the role of teacher and student gender and teacher age on the associations between teacher-rated temperament, EC, and school grades in Math and ML. It was hypothesized that (a) compared with other dimensions, the components of EC and traits related to task orientation would be associated with student school grades most strongly; (b) boys' temperament, EC and school grades would be rated significantly lower than those of girls independent of teacher gender; (c) male teachers would rate student temperament, EC, and school grades higher than female teachers independent of student gender; (d) older teachers would evaluate student temperament and EC more negatively, and give lower grades independent of

student and/or teacher gender; and (e) because of subject-specific demands, teacher-perceived temperament and particularly traits related to task orientation would be more clearly associated students' Math grades than ML grades.

Study IV

Study IV examined whether teacher-rated temperament is more strongly associated with student school grades in Math and ML than self-rated temperament. This was done by including student self-rated temperament and teacher-rated temperament in the same model and therefore controlling for self-rated temperament. In addition, a factor analysis was used to examine whether the factor structure of the six examined temperament traits depended on the rater. Based on the results of the factor analysis, all subsequent analyses were run with both the composition of teachability and six different temperament traits to investigate whether the results would replicate both constructs and to maintain as detailed and trait-specific temperament information as possible. It was hypothesized that (a) teacher-rated temperament is a stronger factor in student school grades than a student's self-rated temperament, even after controlling for student self-rated temperament. This would be true (b) in relation to all temperament traits, and (c) in relation to both subjects grades. Furthermore, it was hypothesized that the factor structure of teacher-rated temperament is similar to that of student self-rated temperament.

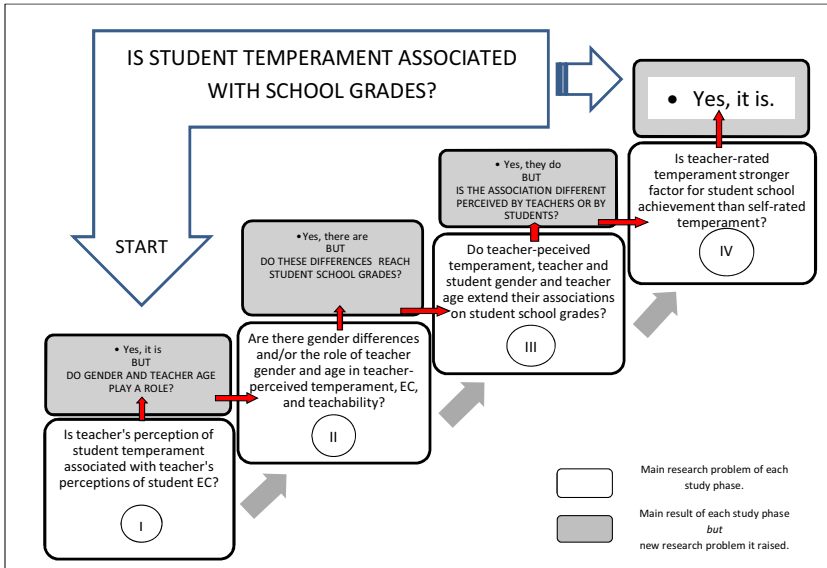


Figure 4. The process and study steps of the research. Each study solved the question set for it, *but* raised new research problems as well. Unsolved and new research problems functioned as a building block and starting point for the next study.

Note. Although the progress of the study steps is like presented in Figure 4, the publication process of four articles has been emerged in different order due to the publication schedule of the journal in question. The dates when articles have been sent and received to the different journals are as follows: Study I / 18th December 2008; Study II / 19th March 2010; Study III / 1st December 2010; and Study IV / 16th September 2011. The article histories are on display in the original articles.

3 Methods

3.1 Design of the study and selection of participants

3.1.1 Design of the Finnish Study of Temperament and School Achievement (FTSA)

The frame of reference for the current study is a wide national research project entitled the *Finnish Study of Temperament and School Achievement* (FTSA) launched in Finland in 2004. The FTSA was designed to study and recognize the important markers related to students' characteristics and school structures which could be factors in students' possible early exclusion.

The main objectives of the FTSA have been to examine the associations of student characteristics (e.g., temperament, self-esteem), teacher perceptions (e.g., ideal or difficult student), teacher-student interaction, and school structures (e.g., class and school size) with student school performance (e.g., school grades) and with student psycho-social well-being (e.g., self-concept)³. The current study is one of the sub-studies conducted on the basis of the FTSA data.

3.1.2 Description of the sample, and sample collection in the FTSA study

The *FTSA* is a geographically representative sample of Finnish upper-comprehensive schools. The population of the sampling was the total number of Finnish upper-comprehensive school students in 2004 ($N = 192\,459$). The study was carried out between autumn 2005 and spring 2006 when the number of ninth graders in Finnish upper-comprehensive school students was 65 137.

The sampling frame used provinces and the number of the schools in each, in which the number of selected schools was proportioned. First, Finland was divided into 5 provinces with a total of 636 schools. All schools in these areas were first listed, and then from each province 10% of the Finnish-speaking schools were randomly selected (Province of Lapland 7%, Province of Oulu 10%, Province of East Finland 12%, Province of West Finland 35% and Province of South Finland 36%). Second, the upper-comprehensive schools' division into rural schools and town schools was computed within each province. The population of each province as a percentage of the popu-

³ The general design of the FTSA study can be obtained from the author.

lation of the whole country was used as an extra criterion for the final sampling.

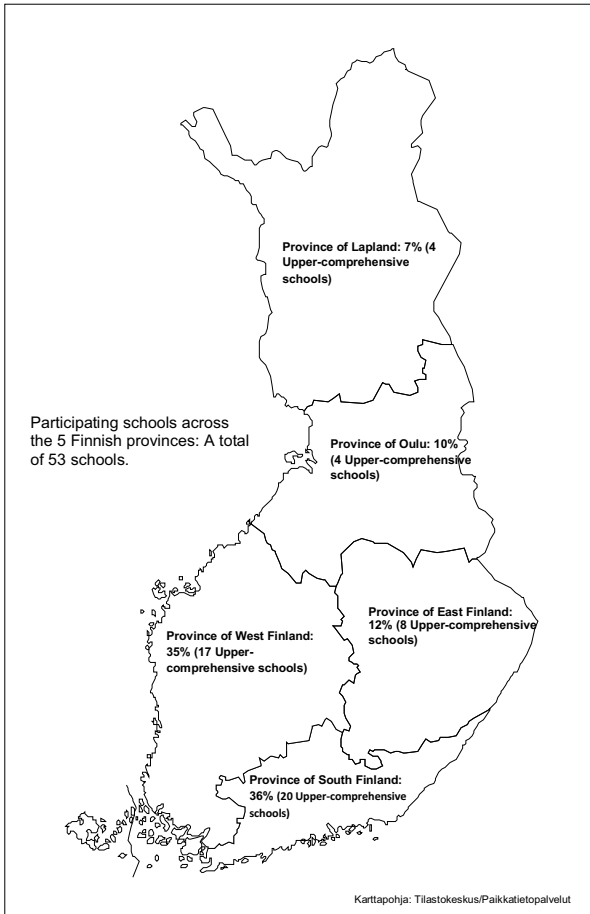


Figure 5. Sampling frame of the study.

The study employed three-phase systematic random sampling, using a k^{th} method, in which the starting point is chosen at random and thereafter at regular intervals. The most common form of systematic sampling is an equal-probability method, where every k^{th} element in the frame is selected, and where k , the sampling interval (sometimes known as the *skip*), is calculated as $k = N / n$ (where n is the sample size, and N is the population size).

The original sample consisted of 64 schools ($N=5992$), from which 43 (67%) agreed to the request. If a school refused to participate, the next k^{th} school on the list was randomly selected resulting in 10 more schools. As a result, the sample consisted of 3901 students in 53 schools across the 5 Finnish provinces (see Figure 5). Swedish-speaking and special schools were not sampled.

Additionally, one teacher from each school was asked to participate as an additional rater. As a result, there were 274 teachers (75% females, 25% males) assessing for 4010 students (94% of all participants) in the final sample.

3.1.3 Participants in the present study

The characteristics of the data used in studies I–IV are given in Table 1. The participants of studies I and II included 3212 ninth grade adolescents (1619 girls and 1593 boys) and 221 teachers (166 women and 55 men) (27 ML teachers, 43 Math teachers, and 151 mainly homeroom teachers or other subject teachers, who had taught the students the longest and thus knew them best).

For studies III and IV, the data on student-teacher pairs for Math and ML were obtained from 1079 ninth grade students and 73 teachers. As there was only one male teacher in the ML teacher group, he and his 16 students were excluded from the analysis. As a result, the final sample of studies III and IV consisted of 1063 ninth grade students (529 girls and 534 boys) participating in the classes of 43 Math teachers (26 females and 17 males) and 29 ML teachers (all females). There were no special teaching groups or bias for any exceptional reasons (e.g., for special education or for gifted and talented students). Ninth graders were selected as subjects because ninth grade is the final year of compulsory schooling, and when they receive the final basic education assessment; it is therefore a significant transition phase along a student's educational career. On the basis of the ninth grade assessment, students will apply and be selected for future studies following comprehensive school.

Table 1. Characteristics of the Data Used in Studies I–IV

Participants	Studies I and II			Studies III and IV		
	Number	%	Age <i>M</i> (<i>sd</i>)	Number	%	Age <i>M</i> (<i>sd</i>)
<i>Students</i>	3212		15.1 (0.37)	1063		15.1 (0.33)
Girls	1619	50.4		529	49.8	
Boys	1593	49.6		534	50.2	
<i>Teachers</i>	221		46.1 (9.53)	72		
Women	166	75.1		55	76.4	47.0 (10.33)
Men	55	24.9		17	23.6	50.6 (7.88)

3.1.4 Procedure of the present study

The municipal board of education and culture gave verbal informed consent to participate. In each selected school, the principal made the decision to participate and, within each of the classes, the decision to participate was made by the homeroom teacher. Given that the students were minors, written informed consent was obtained from their parents.

The teacher ratings were obtained between October 2005 and February 2006, and in compliance with three conditions: (a) that the same Math or ML teacher responsible for giving the Math or ML grade for a student also rated the student's temperament and EC, (b) that each student was rated by one teacher, and (c) that the student had been attending the same class taught by the same teacher for more than one year.

The teacher and student provided his/her ratings of the student's temperament and EC by completing a test battery using paper and pencil. This was done voluntarily and without payment over a 1.5-hour period (2×45 min) during a normal school day. The number of students rated by each teacher varied from 1 to 31 (mean 14.5 rated by one teacher) in studies I and II and from 1 to 25 (mean 14.8 rated by one teacher) in studies III and IV⁴. All measures were administered in the same order without randomization and at the same time in one test session, and there were no systematic reasons for any dropout of students. It was not required that students be rated by both a male and a female teacher. The number of participants varied slightly across the analyses due to some uncompleted responses.

⁴ Due to the multilevel nature of the data (within students in Level 1 and between teachers in Level 2) and the clustering of observations, hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002; Singer & Willett, 2003) was applied for the main analysis in Studies II, III, and IV.

3.2 Measures

The study variables and main statistical methods used in studies I–IV are given in Table 2.

3.2.1 Teacher-perceived student temperament (Study I, II, III, and IV)

Teacher-rated student temperament was assessed using four age-appropriately formed scales from the Temperament Assessment Battery for Children—Revised (TABC-R) (Martin & Bridger, 1999) and two scales from the Revised Dimensions of Temperament Survey (DOTS-R) (Windle, 1992; Windle, Hooker, Lerner, East, Lerner, & Lerner, 1986; Windle & Lerner, 1986) altogether comprising 41 items rated on a 5-point scale ranging from one (*strongly disagree*) to five (*strongly agree*). The temperament dimensions addressed in the TABC-R were activity (four items, e.g., “The pupil seems to have difficulty sitting still”), persistence (eight items, e.g., “A pupil is capable of keeping on with his/her activities for a long time”), inhibition (nine items, e.g., “The pupil takes a long time to become comfortable in a new situation”), and negative emotionality (eight items, e.g., “The pupil lets other students know when he/she does not like something by yelling and fighting”). Martin and Bridger (1999) have reported fairly strong internal consistency reliability for TABC-R dimensions. Cronbach’s alpha reliabilities for activity, persistence, inhibition and negative emotionality have been .86, .93, .87 and .90, respectively, for teacher-reported temperament, and .71, .76, .84 and .82, respectively, for parent-reported temperament (Martin & Bridger, 1999), indicating higher coefficients for Teacher Form in comparison with Parent Form. Previous research has also shown support for the predictive validity of teacher-rated TABC-R in predicting first-grade readiness (Schoen & Nagle, 1994) and off-task behaviour (Orth & Martin, 1994) among kindergarten children, and social status (Hintsanen et al., 2010) and Math achievement (Hintsanen et al., 2012) among adolescents.

The temperament dimensions addressed by the DOTS-R were mood (seven items concerning the tendency to frequently experience a positive feeling, and the amount of pleasant and friendly behaviour in various situations) and distractibility (five items concerning the tendency to be able to concentrate and maintain perceptual focus despite extraneous stimuli). The internal consistency and Cronbach’s alpha reliability coefficients of the DOTS-R factors has been shown to be highly satisfactory across different dimensions, assessments and samples (Guerin et al., 2003; Slabach, Morrow & Wachs, 1991; Windle, 1992) ranging from .54 to .81 in a sample of elementary students (Windle & Lerner, 1986) and .61 to .90 in a sample of ado-

lescent self-reports (Guerin et al., 2003). Specifically for mood and distractibility dimensions, the coefficient alphas have previously been .89 and .81, respectively, for a sample of young adults (Windle & Lerner, 1986), .91 and .79, respectively, for teenagers' self-reports (Windle, 1992) and .87 and .83, respectively, for a sample of adolescents' self-reports (Guerin et al., 2003). Windle and colleagues (Windle, Hooker, Lerner, East, Lerner, & Lerner, 1986) have also reported results which provide initial support for the predictive validity of the DOTS-R for early and late adolescents' perceived cognitive competence, social competence, and general self-worth.

Table 2. Research Design, Study Variables, and Main Statistical Method Used in Studies I-IV

	Study I	Study II	Study III	Study IV
Research design	Cross-sectional	Cross-sectional	Cross-sectional	Cross-sectional
Study variables and measures				
Independent variables	Teacher-perceived temperament (TABC-R & DOTS-R) and EC	Student gender Teacher gender Teacher age	Teacher-perceived temperament (TABC-R & DOTS-R) and EC	Teacher-rated and self-rated temperament (TABC-R & DOTS-R)
Dependent variables	ML grade Math grade	Teacher-perceived temperament (TABC-R & DOTS-R) and EC	ML grade Math grade	ML grade Math grade
Control variables	EC Student gender	Student gender Teacher gender Teacher age	Student gender Teacher Gender Teacher age	Student gender Teacher gender Teacher age
Main statistical method	Hierarchical Regression analysis	Multi-level modelling	Multi-level modelling	Factor analysis Multi-level modelling

Note. EC = Educational competence (i.e., cognitive ability, motivation and maturity developed for the study). ML = Mother language.

In the current study, the Cronbach's alpha reliabilities for the teacher-rated scales were .86, .93, .90, .81, .96, and .91 for activity, persistence, inhibition, negative emotionality, mood, and distractibility, respectively. After statistical

standardization of the variables, sum scores were calculated for each dimension.

3.2.2 Student self-rated temperament (Study IV)

Student self-rated temperament was assessed using four age-appropriately formed scales from the Temperament Assessment Battery for Children—Revised (TABC-R) (Martin & Bridger, 1999) and two scales from the Revised Dimensions of Temperament Survey (DOTS-R) (Windle, 1992; Windle et al., 1986; Windle & Lerner, 1986), comprising a total of 39 items. All items were rated on a five-point scale ranging from one (*strongly disagree*) to five (*strongly agree*). The temperament traits addressed in the TABC-R were activity (six items concerning tendency to engage in energetic gross motor activity), persistence (five items concerning tendency to continue attempting to solve difficult learning or performance problems), inhibition (eight items concerning the tendency to physically withdraw or become emotionally upset when in a social situation which involving persons not previously known), and negative emotionality (eight items concerning tendency to become emotionally upset and express negative emotionality). The temperament traits addressed by the DOTS-R were as follows: positive mood (seven items concerning tendency to frequently experience a positive feeling, and the amount of pleasant and friendly behaviour in various situations) and distractibility (five items concerning tendency to be able to concentrate and maintain perceptual focus despite extraneous stimuli).

For the current sample ($N = 1063$), the Cronbach's alpha reliabilities for self-reported activity, persistence, inhibition, negative emotionality, positive mood, and distractibility were as follows: .51, .60, .84, .68, .91, and .71, respectively. Following statistical standardization of the variables, sum scores were calculated for each dimension.

3.2.3 Teachability (Study IV)

Both teacher-rated and self-rated student *teachability* were assessed using the results of the factor analysis examined for the current data (see Table 9)⁵. For subsequent analyses of teachability, the factor structure of self-rated temperament was applied, which was also consistent with the three-factor construct of teachability introduced by Keogh and associates (Keogh, 1994, 2003). Three teachability scales were formed to analyse both teacher- and

⁵ In Studies I, II, and III the concept of 'Teachability' has been used with the original factor structure from Keogh's theory. In Study IV the concept of 'Teachability' has been assessed using the results of the factor analysis examined for the current data.

self-rated teachability: (1) task orientation (composed of the temperament traits of persistence and distractibility); (2) reactivity (composed of the temperament traits of negative emotionality and activity); and (3) personal-social flexibility (composed of the temperament traits of inhibition and mood).

The Cronbach's alpha reliabilities of the current data for task orientation, reactivity and personal-social flexibility were as follows: .96, .87 and .94, respectively, for teacher-reported teachability and .73, .67 and .87, respectively, for self-rated teachability. Following statistical standardization of the variables, sum scores were calculated for each teachability factor.

3.2.4 Educational competence (Study I, II, and III)

Teacher-rated student *educational competence* (EC) was assessed with three scales developed for this study, and covered cognitive ability (two items, i.e., "The pupil has an equal or higher capacity for handling theoretical studies compared with pupils of the same age", "The pupil has a lower capacity for handling theoretical studies compared with pupils of the same age" [reverse scored]), motivation (four items, i.e., "The pupil is diligent/hardworking", "The pupil would obtain better grades if he/she tried harder" [reverse scored]) and maturity (two items, i.e., "The pupil is well-adjusted compared with other students of the same age", "The pupil is not as well-adjusted as other students of the same age" [reverse scored]). All items were rated on a five-point scale ranging from one (strongly disagree) to five (strongly agree). The Cronbach's alpha reliabilities for the scales were 0.89, 0.90 and 0.86 for cognitive ability, motivation and maturity, respectively. Because the intercorrelations among these three scales were rather high, ranging from 0.60 to 0.67, a global EC score was also derived by adding the three scale scores together. The Cronbach's alpha reliability for the EC scale was 0.92. The factor analysis with Maximum-Likelihood extraction and oblimin rotation also supported the one-factor solution. These procedures were followed in order to reduce possible multicollinearity between the study variables.

3.2.5 School grades (Study I, III, and IV)

The respective grades were taken from the students' latest school reports for Math ($M = 7.50-7.70$, $SD = 1.34-1.35$ and $M = 7.52$, $SD = 1.33-1.41$, respectively, for girls and boys; $n = 1063-2312$) and ML ($M = 8.31-8.33$, $SD = 0.97-0.99$ and $M = 7.46-7.57$, $SD = 1.08-1.10$, respectively, for girls and boys; $n = 1063-2312$); (range = 4-10; 4 means fail, 5 adequate, 6 moderate, 7 satisfactory, 8 good, 9 very good and 10 excellent knowledge and skills). For the

purposes of the study, the teacher who subject-dependently assigned the school grade either in Math or in ML also rated the student's temperament.

3.3 Statistical analyses

Study I

First, gender differences with two-way (gender x temperament and gender x EC) and three-way (gender x temperament x EC) interactions in relation to school grades were tested. Second, a series of linear hierarchical regression analyses were conducted to assess the multivariate relationship between the temperament dimensions and ML and Math grades, with either one as a continuous dependent variable and the temperament dimensions as continuous independent variables. Third, EC was added to the model in order to assess its possible mediating role in the relationship between the temperament dimensions and the two school grades. Model 1 included gender and the six temperament dimensions. Model 2 included gender, the temperament dimensions, and EC. A mediating statistical association was supported if the association between temperament and the ML or Math grade was significantly attenuated after controlling for EC (Baron & Kenny, 1986). A Sobel test (Baron & Kenny, 1986; Sobel, 1982) was also used to confirm the statistical significance of the mediating EC association.

Finally, the interactive associations between EC and the temperament dimensions were tested by linear regression analysis, using gender-adjusted centralized values. Independent variables were entered in four steps in the following order: (a) gender; (b) the predictor variable (temperament dimension); (c) the moderator variable (EC); and (d) the predictor x moderator interaction term.

Study II

Hierarchical linear modeling (HLM) (Boyle & Willms, 2001; Bryk & Raudenbush, 1992; Singer, 1998; Singer & Willett, 2003) was used for the primary tests mainly for two reasons: (1) the data had a natural multi-level data structure comprised of two levels of nesting (within students in Level 1 and between teachers in Level 2), and (2) it was hypothesized that an intra-class correlation coefficient (ICC) would exist between the single observations of students and certain teachers.

First, ICCs were calculated for each of 10 outcome variables from unconditional models, which indicated significant variance in temperament dimension means and in EC means between students taught by different teachers. Second, a Level 1 fixed factor (the student's dummy-coded gender) was

added to the model. Third, a Level 2 fixed factor (the teacher's dummy-coded gender) and Level 2 covariate (the teacher's grand-mean-centred age) were added to the model to quantify the influence of teachers' gender and age on students' temperament and EC means. Finally, a fully adjusted random slope model with one Level 1 fixed factor (Gender-S), one Level 2 fixed factor (Gender-T) and one Level 2 covariate (Age-T) was reset to examine the main associations and interactive associations of Gender-S, Gender-T, and Age-T with teacher-perceived student temperament and EC. Pseudo- R^2 effect sizes were calculated to indicate the percentage of proportion reduction of unexplained variance in each variance component (VC) at each level (Raudenbush & Bryk, 2002; Singer & Willett, 2003).

Study III

Hierarchical linear modeling (HLM) (Raudenbush & Bryk, 2002; Singer & Willett, 2003) was used to take into account the clustering of observations. It calculates the standard errors of the estimates correctly, and allows for a simultaneous examination of both individual and group independent variables (Raudenbush & Bryk, 2002)

In a preliminary step, an unconditional model (Model 0) was fitted for both Math and ML grades without explanatory variables to calculate the intraclass correlation which indicates the proportion of total variance accounted for a between-group variance (compared to variance between individual observations). Following the unconditional model, a random-intercept model was calculated. Student gender was added to the model, after which teacher gender and student temperament and EC traits (all grand-mean-centered, entered separately into the model one at a time) and teacher age (Age-T, grand-mean-centered) were added as well to estimate the role of teacher gender and age and the associations of student temperament and EC with Math and ML grades. The random-intercept model was then extended into a random coefficient model, in which student gender and temperament and EC traits (entered separately into the model one at a time) were allowed to vary over teachers, after which interaction associations between student characteristics and teacher gender and age were tested. This assessed whether the associations between student characteristics and school achievement varied depending on the teacher, and whether teacher gender and age explained any of this variance. The selection of the final model and covariance structure was based on Akaike's (AIC) and Schwarz's Bayesian (BIC) information criteria in order to maximize the number of significant covariance parameters. Based on this procedure, the random-intercept model was best

fitted to the data and chosen for the final models and covariance structures for all further analysis.

Study IV

Two separate factor analysis models were fitted to compare the factor structures of six teacher-rated and student-rated temperament traits to determine whether the factor structure was dependent on the rater. The factorial extraction method was the Principal Axis Factoring (PAF) with Varimax-rotation for six teacher- and self-rated temperament traits, extracting the number of factors with an Eigenvalue greater than one.

Based on the factor analyses, three teachability scales were formed to analyse both teacher- and self-rated teachability: (1) task orientation (persistence and distractibility), (2) reactivity (negative emotionality and activity) and (3) personal-social flexibility (inhibition and mood). All subsequent analyses were run parallel with both the teachability construct and six different temperament scales. This was done for both teacher- and self-rated temperaments to ensure the most detailed trait-specific information regarding the results and to compare whether the results would replicate with both structures.

Again, because of the clustering of observations, a hierarchical linear modeling (HLM) (Raudenbush & Bryk, 2002; Singer & Willett, 2003) was used which calculates the standard errors of the estimates correctly, and allows for a simultaneous examination of both individual and group independent variables (Raudenbush & Bryk, 2002).

As a preliminary step, an unconditional model without explanatory variables was fitted to calculate the intraclass correlation, which indicates the proportion of total variance accounted for by the between-group variance (as compared to variance between individual observations). Following the unconditional model, a two-step analysing procedure was adopted. First, separate random-intercept multilevel linear regression models were calculated to estimate the association of the Math and ML grades with teacher-rated (one set of univariate models) and self-rated (another set of univariate models) temperament (all grand-mean-centred, entered separately as a covariate into the model, one at a time), after adjusting for student and teacher gender (Level 1 and 2 fixed factors; both dummy-coded) and teacher age (Level 2 covariate; grand-mean-centered). The Pseudo- R^2 effect sizes were then calculated to indicate the percentage of the proportion reduction of the unexplained variance in each variance component (VC) at each level (Raudenbush & Bryk, 2002; Singer & Willett, 2003).

Second, to examine whether teacher-rated and student-rated temperament traits had independent associations with school achievement, the random

intercept model was extended by adding teacher-rated and self-rated temperament to the same model (all grand-mean-centred, entered concurrently into the model, one trait at a time) to estimate the mutually adjusted teacher-rated and self-rated temperament associations with Math and ML grades, when both student and teacher gender as well as teacher age were controlled for.

4 Results

The main results of the original four studies are summarized below. Details are given in the original publications (see articles I, II, III, and IV).

4.1 Teachers' perceptions of student temperament, educational competence, and teachability

4.1.1 Main associations of teacher and student gender, and teacher age with teacher-perceived temperament, educational competence, and teachability (Study II)

The results of the multi-level modelling analyses for the associations of teacher gender (Gender-T), student gender (Gender-S), and teacher age (Age-T) with teacher-perceived student temperament and EC are given in Table 3A and 3B.

The main associations of both teacher gender and student gender were statistically significant for activity ($B = 0.216$, $p = .016$ and $B = 0.507$, $p < .001$, respectively), persistence ($B = -0.228$, $p = .001$ and $B = -0.427$, $p < .001$, respectively), and negative emotionality ($B = 0.252$, $p = .006$ and $B = 0.262$, $p < .001$, respectively). The main associations of teacher gender and student gender were also significant for EC ($B = -0.291$, $p = .001$, $B = -0.590$, $p < .001$, respectively), including cognitive ability ($B = -0.393$, $p < .001$, $B = -0.396$, $p < .001$, respectively), motivation ($B = -0.259$, $p = .021$ and $B = -0.702$, $p < .001$, respectively), and maturity ($B = -0.255$, $p = .021$; $B = -0.567$, $p < .001$, and $B = 0.007$, $p = .008$, respectively for teacher gender, student gender, and teacher age).

The results indicate that male teachers rated girls' activity and negative emotionality significantly higher, but persistence and EC significantly lower, in comparison with female teachers' ratings of girls. Independent of their gender, teachers rated boys significantly higher in activity and negative emotionality, but significantly lower in persistence and EC. In addition, the main association of student gender was statistically significant for distractibility ($B = 0.538$, $p < .001$), inhibition ($B = 0.150$, $p = .010$), and mood ($B = -0.209$, $p < .001$). This indicates that teachers have generally perceived boys as significantly higher in distractibility and inhibition, but significantly lower in mood, in comparison with girls.

Table 3A. Summary of the main associations and interactive associations of teacher gender (Gender-T), student gender (Gender-S) and teacher age (Age-T) with teacher-perceived temperament.

Parameter	<i>B</i>	<i>SE</i>	<i>p</i> -value	<i>Pseudo-R</i> ²
<i>Activity</i>				
Intercept	2.132	0.035	<0.001	
Gender-T	0.216	0.072	0.016	0.056 ^b
Gender-S	0.507	0.044	<0.001	0.090 ^a
<i>Persistence</i>				
Intercept	3.794	0.026	<0.001	
Gender-T	-0.228	0.053	0.001	0.087 ^b
Gender-S	-0.427	0.031	<0.001	0.114 ^a
Gender-T x Gender-S	0.163	0.064	0.012	0.130 ^c
<i>Distractibility</i>				
Intercept	2.641	0.034	<0.001	
Gender-T	0.156	0.069	0.300ns	0.013 ^b
Gender-S	0.538	0.039	<0.001	0.116 ^a
Gender-T x Gender-S	-0.195	0.082	0.018	0.119 ^c
<i>Inhibition</i>				
Intercept	2.669	0.026	<0.001	
Gender-T	0.148	0.053	0.109ns	0.020 ^b
Gender-S	0.150	0.027	0.010	0.022 ^a
Gender-T x Gender-S	-0.152	0.057	0.009	0.111 ^c
Gender-T x Gender-S x Age-T	0.012	0.006	0.049	0.125 ^c
<i>Negative Emotionality</i>				
Intercept	2.174	0.036	<0.001	
Gender-T	0.252	0.074	0.006	0.072 ^b
Gender-S	0.262	0.038	<0.001	0.047 ^a
<i>Mood</i>				
Intercept	3.671	0.037	<0.001	
Gender-T	-0.120	0.075	0.236ns	0.009 ^b
Gender-S	-0.209	0.039	<0.001	0.041 ^a
Gender-T x Gender-S x Age-T	-0.020	0.009	0.022	0.100 ^c

Note. *SE* = standard error of estimate; Gender-T = gender of teachers (dummy-coded); Gender-S = gender of students (dummy-coded); Females and girls serve as the reference category. Age-T = teachers' centered age. All single associations of Gender-T and Gender-S are reported. Otherwise, only statistically significant findings of the fully adjusted random slope model are reported. All results have Bonferroni adjustment for multiple comparisons. *ns* = Nonsignificant. *Pseudo-R*² = percentage of reduction of unexplained variance of the unique variable; Baselines for model

comparisons: ^a = residual variance, ^b = intercept variance, ^c = random slope variance. Covariance parameters (all *p-values*<.01) are omitted from the table (presented in original article).

Table 3B. Summary of the main associations and interactive associations of teacher gender (Gender-T), student gender (Gender-S) and teacher age (Age-T) with teacher-perceived Educational Competence.

Parameter	<i>B</i>	<i>SE</i>	<i>p-value</i>	<i>Pseudo-R</i> ²
<i>Educational Competence</i>				
Intercept	3.857	0.031	<0.001	
Gender-T	-0.291	0.063	0.001	0.080 ^b
Gender-S	-0.590	0.039	<0.001	0.114 ^a
Gender-T x Gender-S	0.239	0.080	0.003	0.172 ^c
<i>Cognitive Ability</i>				
Intercept	3.948	0.043	<0.001	
Gender-T	-0.393	0.087	<0.001	0.123 ^b
Gender-S	-0.396	0.049	<0.001	0.040 ^a
Gender-T x Gender-S	0.204	0.103	0.048	0.121 ^c
<i>Motivation</i>				
Intercept	3.766	0.034	<0.001	
Gender-T	-0.259	0.071	0.021	0.035 ^b
Gender-S	-0.702	0.043	<0.001	0.134 ^a
Gender-T x Gender-S	0.262	0.090	0.004	0.130 ^c
<i>Maturity</i>				
Intercept	3.946	0.034	<0.001	
Gender-T	-0.255	0.069	0.021	0.049 ^b
Gender-S	-0.567	0.037	<0.001	0.088 ^a
Age-T	0.007	0.004	0.008	0.082 ^b
Gender-T x Gender-S	0.243	0.077	0.002	0.013 ^a

Note. *SE* = standard error of estimate; Gender-T = gender of teachers (dummy-coded); Gender-S = gender of students (dummy-coded); Females and girls serve as the reference category. Age-T = teachers' centered age. All single associations of Gender-T and Gender-S are reported. Otherwise, only statistically significant findings of the fully adjusted random slope model are reported. All results have Bonferroni adjustment for multiple comparisons. *ns* = Nonsignificant. *Pseudo-R*² = percentage of reduction of unexplained variance of the unique variable; Baselines for model comparisons: ^a = residual variance, ^b = intercept variance, ^c = random slope variance. Covariance parameters (all *p-values*<.01) are omitted from the table (presented in original article).

4.1.2 Interactive associations of teacher and student gender, and teacher age with teacher-perceived temperament, educational competence, and teachability (Study II)

Table 3A and Table 3B also show statistically significant two-way teacher gender \times student gender interactions in relation to persistence ($B = 0.163, p = .012$), distractibility ($B = -0.195, p = .018$), inhibition ($B = -0.152, p = .009$), and EC ($B = 0.239, p = .003$) (including cognitive ability $B = 0.204, p = .048$, motivation $B = 0.262, p = .004$, and maturity $B = 0.243, p = .002$). In addition, significant three-way teacher gender \times student gender \times teacher age interaction is evident in relation to inhibition ($B = 0.012, p = .049$) and mood ($B = -0.020, p = .022$). The two-way interactions, illustrated in Figure 6, suggest that the gender gap between male teachers' ratings for boys and female teachers' ratings for girls was lower in persistence, EC, distractibility, and inhibition than could be concluded merely from the main associations of teacher and student gender. This means that male teachers perceived boys' and girls' persistence, EC, distractibility, and inhibition as closer to each other than female teachers did. However, inhibition was higher and mood lower for boys as assessed by older male teachers compared with girls as assessed by a female teacher.

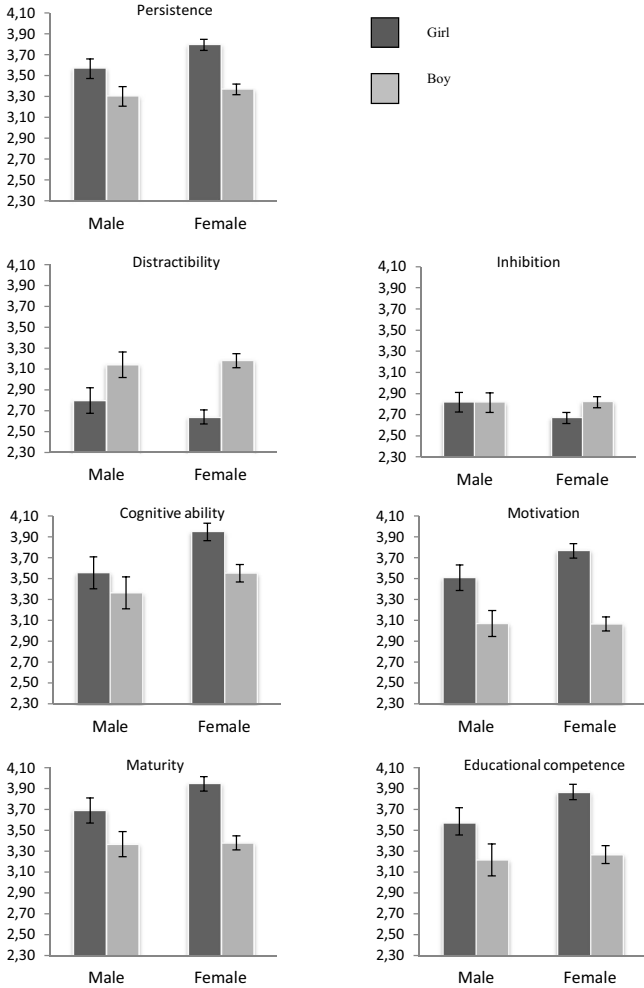


Figure 6. Interactive associations of teacher gender (Gender-T) and student gender (Gender-S) with teacher-perceived temperament, educational competence (EC; including the dimensions of cognitive ability, motivation, and maturity) and teachability. Only statistically significant findings are presented ($p < .05$). The vertical bars denote the \pm standard error of the mean (SEM).

4.1.3 Summary of the results of Study II

The four major findings of study II are as follows. First, there was a significant gender difference between girls' and boys' temperament, EC, and teach-

ability. Independent of teachers' gender, girls were evaluated as having higher EC and teachability and were rated higher in temperament traits, reflecting high teachability. Second, there was significant teacher gender x student gender interaction in relation to teacher-perceived persistence, distractibility, inhibition, and EC (including cognitive ability, motivation, and maturity). These associations occurred only with male teachers and both boy and girl students. Third, a significant interactive association was noted between teacher gender, student gender and teacher age, and perceptions of student's temperament, particularly inhibition and mood, occurring only with male teachers and with respect to boys. Fourth, a significant main association was also noted between teacher age and perceptions of student EC, particularly maturity, independent of teacher and student gender.

4.2 Associations of teacher-perceived temperament, educational competence, and teachability with school achievement

4.2.1 Main associations of teacher-perceived temperament and educational competence with teacher gender and age in relation to Mathematics and Mother language grades (Study III)

Main associations of the multilevel modeling analyses for the associations of teacher-perceived temperament, EC, student and teacher gender, and teacher age with students' Math and ML grades are presented in Table 4. Higher activity, distractibility, inhibition and negative emotionality were associated with lower Math and ML grades, with one standard deviation of temperament difference being associated with a -0.26 to -0.58 and -0.21 to -0.50 difference in Math and ML grades respectively. Distractibility was the strongest factor for lower grades for both subjects. Higher persistence, (positive) mood, EC, cognitive ability, motivation and maturity were associated with higher Math and ML grades, with one standard deviation of temperament difference being associated with a 0.20 to 0.77 and 0.16 to 0.61 difference in Math and ML grades, respectively. Persistence and EC were the strongest factors for higher grades for both subjects, whereas (positive) mood had the weakest association for both grades. No main association of teacher gender or teacher age with Math or ML grades was noted.

Table 4. Main associations of teacher and student characteristics with Mathematics and Mother language grades. Separate random-intercept multilevel linear regression models.

	Mathematics		Mother language	
	<i>B (SE)</i>	β	<i>B (SE)</i>	β
<i>Teacher characteristics</i>				
Teacher's male gender	-0.10 (0.17)	-0.08	-	-
Teacher's age	-0.01 (0.01)	-0.01	-0.00 (0.01)	-0.02
<i>Student characteristics</i>				
Student's male gender	-0.01 (0.11)	-0.01	-0.70* (0.10)	-0.64
Activity	-0.61* (0.05)	-0.45	-0.35* (0.04)	-0.32
Persistence	0.83* (0.05)	0.62	0.52* (0.04)	0.47
Inhibition	-0.35* (0.06)	-0.26	-0.23* (0.05)	-0.21
Negative emotionality ^a	-0.50* (0.06)	-0.38	-0.27* (0.05)	-0.24
Mood ^a	0.27* (0.06)	0.20	0.18* (0.04)	0.16
Distractibility	-0.78* (0.05)	-0.58	-0.54* (0.04)	-0.50
Educational competence	1.03* (0.04)	0.77	0.67* (0.04)	0.61
Cognitive ability	0.95* (0.04)	0.71	0.66* (0.04)	0.60
Motivation	0.95* (0.05)	0.71	0.56* (0.04)	0.51
Maturity	0.78* (0.05)	0.58	0.51* (0.04)	0.47

Note. N_s for Math = 26 female teachers, 17 male teachers, and 636 students.

N_s for Mother language = 26 female teachers and 427 students.

B = Unstandardized regression coefficient, SE = standard error,

β (Beta) = Standardized regression coefficient;

All results for temperament and educational competence are presented for standardized scales (Mean=0, $SD=1$), adjusted for student gender.

* $p < 0.001$

Covariance type VC (Variance Components) was chosen for all analyses except those marked with ^a.

^a Covariance type diagonal (ID) in Mother language.

Intercept and covariance parameters (all $ps < .05$) are omitted from the table.

4.2.2 Moderating and mediating associations of educational competence in relation to Mathematics and Mother language grades (Study I)

The results of the linear regression analyses that were performed separately for ML and Math are presented in Table 5. Model 1A and Model 1B indicate the independent contribution of each temperament dimension (adjusted for the other and for gender) to the ML and Math grades, respectively. Activity, persistence, distractibility, inhibition, and negative emotionality were significantly associated with ML and Math grades, explaining 28% and 29% of their variance, respectively.

Table 5. Standardised (β) regression coefficients for teacher-rated temperament dimensions (hierarchically adjusted for gender and educational competence) in relation to student grades in Mother language ($N = 3141$) and Mathematics ($N = 3148$)

Model and variable	β	R^2	ΔR^2	F	df
Mother language					
Model 1A					
Gender	0.207***		0.142	518.22***	1, 3139
Activity	0.098**		0.099	406.85***	1, 3138
Persistence	0.528***		0.031	168.74***	1, 3133
Distractibility	-0.171***		0.036	186.59***	1, 3134
Inhibition	-0.195***		0.112	542.43***	1, 3137
Negative emotionality	0.174***		0.001	3.57	1, 3136
Mood	0.003		0.001	2.47	1, 3135
0.420					
Model 2A					
Gender	0.187***		0.142	518.22***	1, 3139
Activity	0.028		0.099	406.85***	1, 3138
Persistence	0.166***		0.031	168.74***	1, 3133
Distractibility	0.024		0.036	186.59***	1, 3134
Inhibition	-0.110***		0.112	542.43***	1, 3137
Negative emotionality	0.137***		0.001	3.57	1, 3136
Mood	-0.047**		0.001	2.47	1, 3135
+ Educational competence	0.571***		0.084	530.37***	1, 3132
0.504					
Mathematics					
Model 1B					
Gender	-0.102***		0.004	13.77***	1, 3146
Activity	0.199***		0.108	384.46***	1, 3145
Persistence	0.550***		0.034	150.93***	1, 3140
Distractibility	-0.230***		0.052	222.72***	1, 3141
Inhibition	-0.152***		0.094	371.40***	1, 3144
Negative emotionality	0.111***		0.001	2.47	1, 3143
Mood	-0.021		0.000	0.13	1, 3142
0.292					

Table 5 continues

Model 2B				
Gender	-0.123***	0.004	13.77***	1, 3146
Activity	0.128***	0.108	384.46***	1, 3145
Persistence	0.171***	0.034	150.93***	1, 3140
Distractibility	-0.027	0.052	222.72***	1, 3141
Inhibition	-0.063**	0.094	371.40***	1, 3144
Negative emotionality	0.073**	0.001	2.47	1, 3143
Mood	-0.074***	0.000	0.13	1, 3142
+ Educational competence	0.599***	0.093	473.98***	1, 3139
0.385				

Note. The β coefficients are those computed at the final step of each analysis. R^2 is for the whole model. Model 1 = temperament dimensions adjusted for gender. Model 2 = temperament dimensions adjusted for gender and educational competence. * $p < .05$, ** $p < .01$, *** $p < .001$.

Adding EC to the model (Model 2A and Model 2B) resulted in an 8% and 9% increase in R^2 in relation to the ML and Math grades, respectively. Further, the associations of activity and distractibility with ML and distractibility with Math were no longer significant, which provides evidence of mediation.

The Sobel test results (Table 6) (Baron & Kenny, 1986; Sobel, 1982), confirmed that EC was a significant mediator of the statistical associations of all six temperament dimensions in relation to ML and Math grades (all Z 's were significant at the level of $p < .0001$).

Table 6. Mediating results: Effects of temperament on Mother language and Maths grades through Educational competence (EC)

Mediated pathway	DV: Mother language		DV: Mathematics	
	^a Sobel		^a Sobel	
Temperament (IV) --> EC (MV)--> School grade (DV)	Z-value	p-value	Z-value	p-value
IV: Activity	5.27	<.0001	5.25	<.0001
Persistence	16.77	<.0001	16.05	<.0001
Distractibility	-14.85	<.0001	-14.34	<.0001
Inhibition	-10.36	<.0001	-10.18	<.0001
Negative emotionality	4.64	<.0001	4.62	<.0001
Mood	6.57	<.0001	6.52	<.0001

Note. IV = independent variable; MV = mediator variable; DV = dependent variable.

^aSobel Test Results (Sobel, 1982) for Baron & Kenny's (1986) step 4.

Table 7. Interactions between Educational competence and Temperament dimensions (all teacher-rated) in relation to student's grades in Mathematics

Variable	β	R^2	ΔR^2	F	Df
Gender	-0.118***		0.004	13.00***	1, 3156
Activity	0.031		0.108	382.12***	1, 3155
Educational competence	0.663***		0.264	1333.75***	1, 3154
Activity x Educational competence	-0.046**	0.377	0.002	9.47**	1, 3153
Gender	-0.121***		0.004	13.22***	1, 3162
Persistence	0.018		0.224	918.75***	1, 3161
Educational competence	0.634***		0.147	740.92***	1, 3160
Persistence x Educational competence	0.055***	0.377	0.003	13.55***	1, 3159
Gender	-0.119***		0.004	13.18***	1, 3161
Distractibility	-0.005		0.211	849.26***	1, 3160
Educational competence	0.646***		0.160	807.80***	1, 3159
Distractibility x Educational competence	-0.074***	0.379	0.005	26.43***	1, 3158

Note. $N_s = 3158$ (for Activity x Educational competence), 3164 (for Persistence x Educational competence) and 3163 (for Distractibility x Educational competence). The β coefficients are those computed at the final step of each analysis. R^2 is for the whole model. Only statistically significant findings are reported. * $p < .05$, ** $p < .01$, *** $p < .001$.

The regression analysis (Table 7) revealed significant $EC \times$ activity, $EC \times$ distractibility, and $EC \times$ persistence interactions in relation to Math grades ($p=.002$, $p<.001$, and $p<.001$, respectively). These interactions are also depicted in Figure 7, which shows that activity was negatively related to Math grade among students with high EC, but not among students with low EC. However, the associations of persistence and distractibility with Math grades were, in practice, similar on both EC levels.

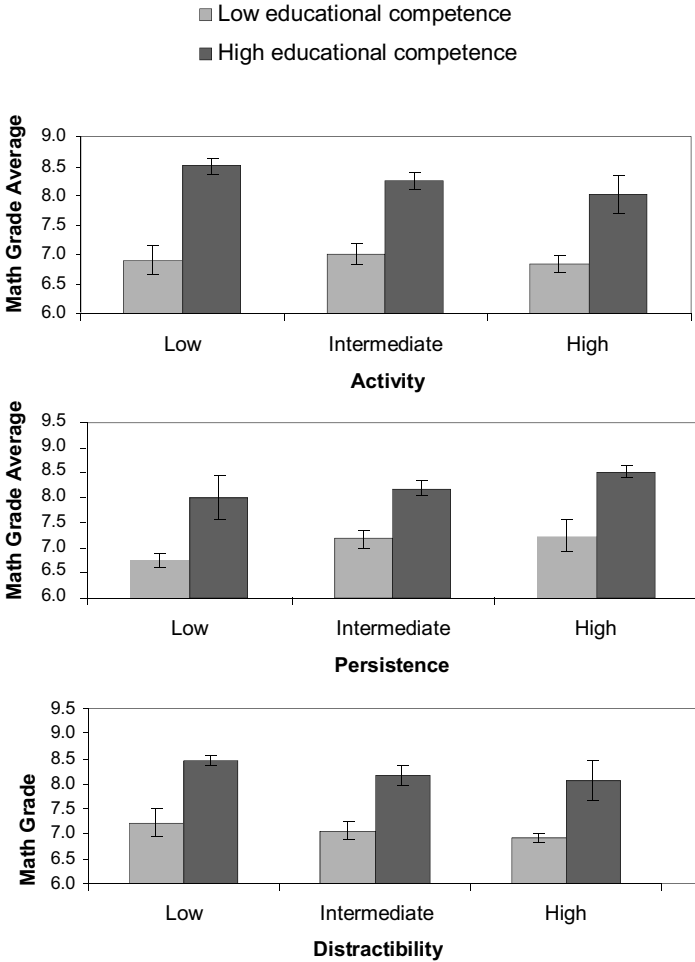


Figure 7. Results of the linear regression analysis estimating the interaction between activity, persistence, distractibility (respectively from top to bottom), and Math grade as a function of low, intermediate and high educational competence (EC). The range of the scale on the x axis is the mean \pm 2 *SD*. The vertical bars denote 99.9% confidence intervals.

4.2.3 Interactive associations of teacher-perceived temperament and educational competence with teacher and student gender, and teacher age in relation to Mathematics and Mother language grades (Study III)

Interactive associations of student temperament and EC with teacher age and student gender in relation to student ML grade are presented in Figure 8. The associations were significant for negative emotionality ($\beta = 0.014, p = 0.005$), EC ($\beta = -0.006, p = 0.054$) and motivation ($\beta = -0.007, p = 0.051$). In addition, interaction between student gender and mood ($\beta = 0.182, p = 0.036$) in relation to ML grades was significant. The narrowing differences between high and low groups by teachers' advancing age indicated that student negative emotionality, EC and motivation were more strongly related to student ML grades among younger teachers than older ones. In addition, student positive mood was associated with higher ML grades more strongly in boys than in girls, as shown by the steeper regression line for boys compared to girls (Figure 8).

Figure 9 presents the gender-specific interactive associations between teacher age and students' inhibition ($\beta = 0.016, p = 0.050$) and maturity ($\beta = -0.018, p = 0.022$) in relation to students' ML grades. These interactive associations indicated that inhibition and maturity were stronger factors in boys' ML grades among older teachers compared to younger ones, while such associations with teachers' age were not observed among girls. No significant interactive associations between teacher gender or age and student characteristics in relation to Math grade were noted.

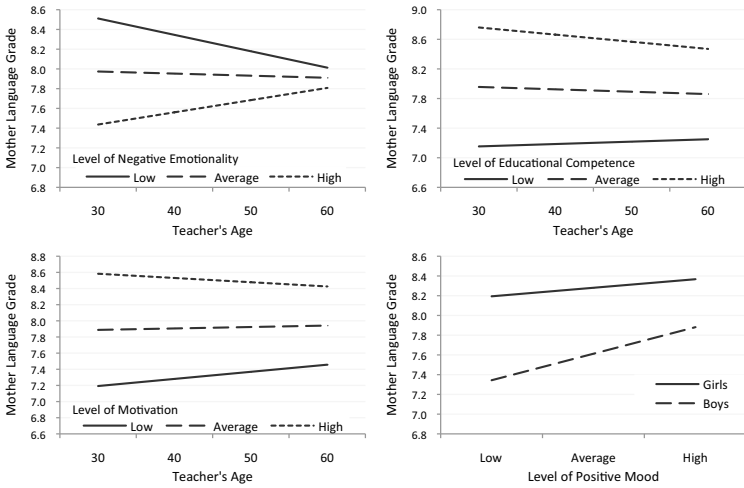


Figure 8. Interactive associations of student temperament with teacher’s age and student gender. Predicted values of Mother language grade are calculated from linear regression models. Low=1 standard deviation below the mean, Average=mean, High=1 standard deviation above the mean.

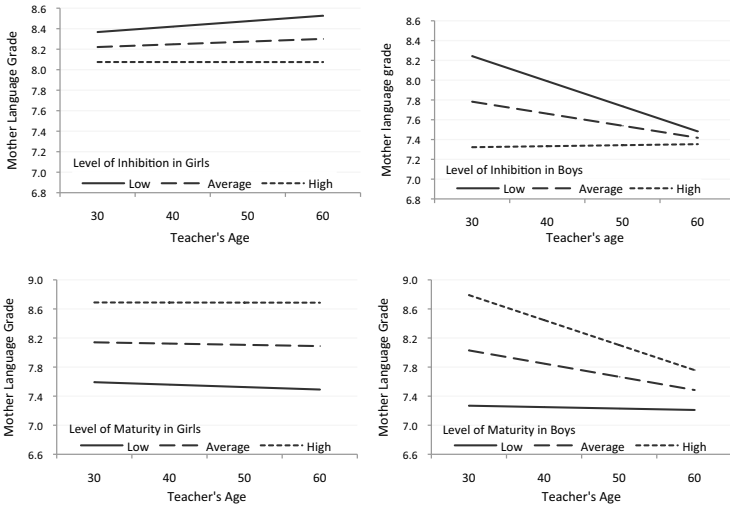


Figure 9. Gender-specific interactive associations of teacher’s age with student’s inhibition (upper panels) and maturity (lower panels). Predicted values of Mother language grade are calculated from linear regression models. Low=1 standard deviation below the mean, Average=mean, High=1 standard deviation above the mean.

4.2.4 Gender differences in Mathematics and Mother language grades (Study III)

Table 8. Gender differences in Mathematics and Mother language (positive values indicate higher grades in boys, negative values indicate higher grades in girls)

Gender difference adjusted for	Mathematics				Mother language			
	β	<i>SE</i>	<i>p</i>	<i>SD</i> difference	β	<i>SE</i>	<i>p</i>	<i>SD</i> difference
None	-0.02	0.11	0.873	-0.01	-0.70	0.10	0.000	-0.63
Activity	0.20	0.10	0.044	0.15	-0.54	0.10	0.000	-0.49
Persistence	0.30	0.09	0.001	0.22	-0.42	0.09	0.000	-0.38
Inhibition	0.04	0.10	0.680	0.03	-0.65	0.10	0.000	-0.59
Negative emotionality	0.12	0.10	0.246	0.09	-0.62	0.10	0.000	-0.56
Mood	0.02	0.10	0.816	0.02	-0.65	0.10	0.000	-0.59
Distractibility	0.28	0.09	0.003	0.21	-0.39	0.09	0.000	-0.35
Educational competence	0.44	0.08	0.000	0.33	-0.28	0.08	0.000	-0.26
Cognitive ability	0.10	0.08	0.178	0.08	-0.47	0.08	0.000	-0.43
Motivation	0.47	0.08	0.000	0.35	-0.35	0.09	0.000	-0.31
Maturity	0.33	0.09	0.000	0.25	-0.38	0.09	0.000	-0.35
All Temperament traits	0.36	0.09	0.000	0.27	-0.26	0.08	0.001	-0.23
All EC traits	0.31	0.08	0.000	0.23	-0.34	0.08	0.000	-0.31
All traits	0.32	0.08	0.000	0.24	-0.27	0.08	0.000	-0.24

Note. β =regression coefficient. *SE*=standard error. *p*=*p*-value for gender difference. *SD* difference=gender difference in expressed in units of standard deviations.

The gender differences in Math and ML grades are presented in Table 8 and also illustrated in Figure 10. Boys had higher Math grades than girls, after being adjusted for activity, persistence, distractibility, EC, motivation, and maturity. In the units of standard deviations, the significant gender difference ranged from 0.15 to 0.21 and from 0.25 to 0.35 after being adjusted for temperament traits and EC traits, respectively, being the strongest after adjustment for motivation. Furthermore, boys received lower ML grades from their female teachers than girls, after adjustment for temperament and EC traits. In units of standard deviations, the significant gender difference ranged from -0.35 to -0.59 and from -0.26 to -0.43 after being adjusted for temperament and EC traits, respectively, being the strongest after adjustment for inhibition

and (positive) mood. However, the difference in ML grades between girls and boys attenuated by 62% after being adjusted for all temperament and EC traits, although girls still had higher ML grades. Generally, the gender differences were stronger in ML than in Math.

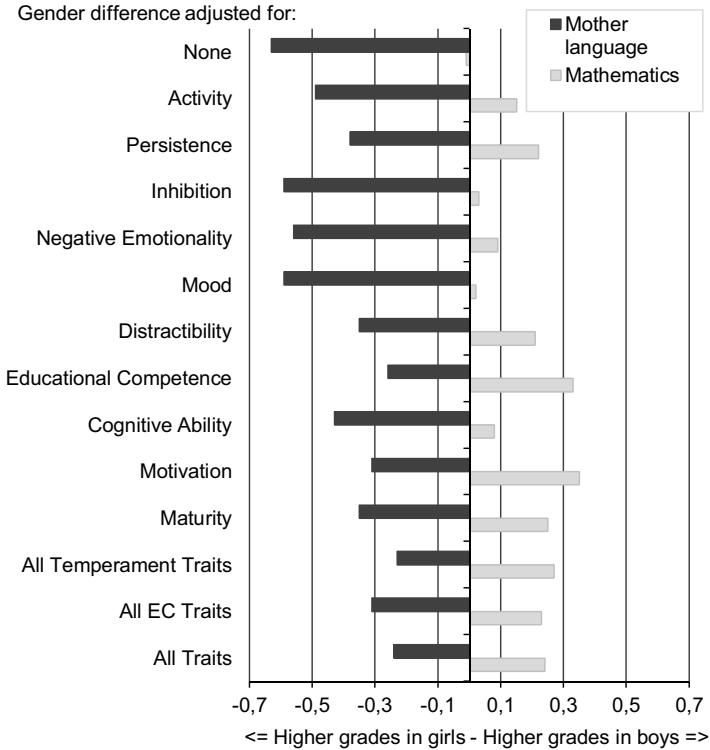


Figure 10. Gender differences (expressed in units of standard deviations of temperament differences being associated with units of standard deviations differences in school grades) in Mother language and Mathematics adjusted for student characteristics (positive values indicate higher grades in boys, negative values indicate higher grades in girls).

4.2.5 Summary of the results of studies I and III

The summary of the results of studies I and III are as follows. First, high temperamental activity, inhibition, negative emotionality and distractibility in particular were associated with lower ML and Math grades, with stronger associations being observed for Math than ML. Second, high temperamental

mood and persistence in particular, and high EC (including cognitive ability, motivation and maturity) were associated with higher school grades, again the associations being stronger for Math than for ML. Third, EC mediated the relationship between temperament and both school grades and moderated the relationship between activity and Math grade. Fourth, boys had lower ML grades but higher Math grades than girls, when teacher-perceived temperament and EC were adjusted for, the gender differences being stronger for ML than for Math.

In addition, interactive associations between student and teacher characteristics were noted as follows. First, positive mood was associated with ML grades more strongly in boys than in girls. Second, teacher age was more strongly associated with lower ML grades in boys compared to girls, independently of teacher-perceived EC, motivation and maturity. Third, inhibition and maturity were stronger factors of boys' (but not girls') ML grades among older teachers compared to younger teachers (inhibition increased and maturity decreased ML grades more markedly in boys than in girls). Fourth, teacher age was shown to be a significant factor; among younger teachers negative emotionality, EC, and motivation were more strongly related to student ML grades than among older teachers.

Comparisons of girls with boys indicated that gender differences in Math grades emerged only when some student characteristics (EC, persistence and distractibility in particular) were adjusted for. On average, girls had higher ML grades than boys, and almost two-thirds (62%) of this difference was accounted for by gender differences in teacher-perceived temperament and EC.

4.3 Teacher-rated versus self-rated student temperament, teachability, and school achievement

4.3.1 Factor structures of teacher-rated versus self-rated temperament (Study IV)

The results of two separate exploratory factor analyses, which were fitted in order to determine the psychometric dimensionality of teacher- and self-rated temperament, are presented in Table 9.

Based on the factor loadings shown in Table 9, the teacher-rated factor structure was comprised of two factors. Factor 1 consisted of the temperament traits of persistence, activity, distractibility, and negative emotionality. Factor 2 consisted of the temperament traits of inhibition and positive mood. The highest loadings focused on persistence (-.99), activity (.91) and distractibility (.85), i.e., traits related to task-orientation.

As shown in Table 9, the self-rated factor structure consisted of three factors. Factor 1 consisted of the temperament traits of positive mood and inhibition. Factor 2 consisted of the temperament traits of persistence and distractibility, and factor 3 consisted of the temperament traits of negative emotionality and activity. The highest loadings focused on persistence (-.63), negative emotionality (.63) and mood (.63), i.e., traits related to task orientation, reactivity, and personal-social flexibility, respectively.

The results indicated that the factor structure of self-rated temperament was more differentiated when compared to that of teacher-rated temperament. In the teacher-ratings, negative emotionality was not differentiated from task orientation, i.e., from the most pervasive factor of 'school temperament', as it was in student self-ratings. The factor structures explained 73% and 44% of the common variance for teacher-rated and self-rated temperament, respectively.

Table 9. Factor loadings of separate factor analysis for teacher-rated and self-rated temperament dimensions

Variable	Teacher-rated			Self-rated			
	F1	F2	Communality	F1	F2	F3	Communality
Persistence	<u>-99</u>	.07	.89	.31	<u>-.63</u>	-.10	.24
Activity	<u>.91</u>	.24	.84	.41	.08	<u>.61</u>	.20
Distractibility	<u>.85</u>	-.05	.74	.02	<u>.60</u>	.06	.15
Negative emotionality	<u>.75</u>	-.05	.60	-.40	.14	<u>.63</u>	.23
Inhibition	-.02	<u>-.78</u>	.45	<u>-.51</u>	.13	.03	.18
Mood	-.02	<u>.78</u>	.43	<u>.63</u>	-.06	.03	.22
Eigenvalue	3.11	1.30		1.01	0.80	0.78	
<i>R</i> ²	51.85	21.57		17.95	13.33	13.04	
<i>Sums of Squared Loadings</i>			73.42				44.32

Note. *Ns* = 1063 ratings for teacher-rated and self-rated temperament.

Extraction Method: Principal Axis Factoring with Varimax-rotation.

Loadings primarily recorded to the certain factor are bold and underlined.

For subsequent analyses, the structure of self-rated temperament was applied, which was shown to be consistent with the three-factor construct of teachability introduced by Keogh and associates (Keogh, 1994; 2003). All subsequent multilevel regression analyses were run parallel with both the teachability constructs and the six different temperament scales. This was done to

ensure the most detailed trait-specific information regarding the results and in order to compare whether the results would replicate with both structures. Both teacher- and self-rated temperament were used, because having multiple raters has been repeatedly recommended in the literature (Carey & McDevitt, 1995; Guerin et al., 2003; Hintsanen et al., 2012; Keogh, 2003; Kristal, 2005; Martin & Bridger, 1999; Rothbart & Jones, 1998), and in order to examine whether the associations were dependent on the rater.

4.3.2 Mutually adjusted associations of teacher- and self-rated temperament and teachability with school grades (Study IV)

The separate multilevel modeling analyses for the trait-specific associations were examined to determine whether teacher-rated temperament would be a stronger factor for school grades than student self-rated temperament. This was examined by including student self-rated temperament and teacher-rated temperament in the same model (mutually adjusted, one trait at a time) and therefore controlling for the association of self-rated temperament.

Table 10A presents the main associations for the trait-specific associations in relation to Math and ML grades in the gender- and age-adjusted models. Teacher-rated higher activity, inhibition, negative emotionality, and distractibility were significantly associated with lower Math and ML grades, with a difference of 1 *SD* in the temperament scores being associated with differences of -0.31 to -0.56 and -0.23 to -0.50 in Math and ML grades, respectively. Likewise, higher teacher-rated persistence and positive mood were significantly related with higher Math and ML grades, with a difference of 1 *SD* in the temperament scores being associated with differences of 0.19 and 0.56, and 0.16 and 0.45 in Math and ML grades, respectively. The associations of teacher-rated temperament with school grades were not attenuated even when the self-ratings of temperament were controlled for, indicating that the associations of the teacher-rated and self-rated temperament with school grades were largely independent of each other⁶.

After an adjustment for teacher-rated temperament, higher self-rated distractibility significantly related with lower Math and ML grades, with a difference of 1 *SD* in the temperament scores being associated with differences of -0.09 and -0.08 in the Math and ML grades, respectively. Likewise, higher self-rated persistence related with higher Math and ML grades, with a difference of 1 *SD* in the temperament scores being associated with differences of 0.15 and 0.18 in Math and ML grades, respectively. However, student self-

⁶ The results of the previous analyses examined separately for teacher- and self-rated temperament and teachability are presented in the original article (Study IV).

rated negative emotionality was associated only with lower ML grades, with a difference of 1 *SD* in the temperament score being associated with a difference of -0.14 in ML grade. Further, the association of self-rated inhibition was statistically significant: higher self-rated inhibition related with higher Math grades with a difference of 1 *SD* in the temperament score being associated with a difference of 0.12 in Math grades.

Table 10A. Associations of teacher-rated and self-rated temperament (concurrently, one trait at a time) with to Mathematics and Mother language grades. Separate random-intercept multilevel linear regression models.

Temperament trait adjusted for student and teacher gender and teacher age	Teacher-rated			Self-rated		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Mathematics						
Activity	-0.45	0.04	<.001	0.04	0.04	0.28
Persistence	0.56	0.04	<.001	0.15	0.03	<.001
Inhibition	-0.31	0.05	<.001	0.12	0.04	0.01
Negative emotionality	-0.39	0.05	<.001	-0.03	0.04	0.49
Positive mood	0.19	0.05	<.001	0.01	0.04	0.80
Distractibility	-0.56	0.04	<.001	-0.09	0.04	0.02
Mother language						
Activity	-0.31	0.04	<.001	0.01	0.05	0.78
Persistence	0.45	0.04	<.001	0.18	0.04	<.001
Inhibition	-0.23	0.05	<.001	0.06	0.05	0.16
Negative emotionality	-0.27	0.04	<.001	-0.14	0.04	0.00
Positive mood	0.16	0.05	<.001	0.01	0.05	0.89
Distractibility	-0.50	0.04	<.001	-0.08	0.04	0.05

Note. *Ns* for Math = 43 teachers and 636 students.

Ns for Mother language = 26 teachers and 427 students.

β = Standardized regression coefficient; All results for temperament and school grades are presented for standardized scales (Mean = 0, *SD* = 1), adjusted for teacher and student gender and teacher age.

Each model includes teacher-rated and self-rated temperament trait.

Covariance type VC (Variance Component) was chosen for all analyses.

Intercept parameters (all *ps* < .001) and covariance parameters (all *ps* < .05) are omitted from the table.

When examining the corresponding analysis with the three components of teachability construct, the results run parallel to and have corresponding effect sizes as the mutually adjusted six temperament traits. Table 10B presents

the main associations of the separate multilevel modeling analyses for the factor-specific associations in relation to Math and ML grades. Both the teacher-rated and self-rated teachability factors were included in the same model (mutually adjusted, one teachability factor at a time) in the gender- and age-adjusted models.

High teacher-rated reactivity (high negative emotionality and high activity) was associated with lower Math and ML grades, with a difference of 1 *SD* in the reactivity scores being associated with differences of -0.54 and -0.32 in Math and ML grades, respectively. Likewise, higher teacher-rated task orientation (high persistence and low distractibility) and high teacher-rated personal-social-flexibility (high positive mood and low inhibition) were significantly associated with higher Math and ML grades, with a difference of 1 *SD* in the teachability scores being associated with differences of 0.24 to 0.58 and 0.20 to 0.51 in Math and ML grades, respectively. The associations between teacher-rated teachability factors and school grades were not attenuated even when the self-ratings of teachability factors were controlled for, indicating that the associations of the teacher-rated and self-rated teachability with school grades were largely independent of each other⁷.

After adjusting for teacher-rated teachability, higher self-rated task orientation (high persistence and low distractibility) was significantly associated with higher Math and ML grades, with a difference of 1 *SD* in the teachability scores being associated with differences of 0.13 and 0.14 in Math and ML grades, respectively. However, student self-rated reactivity (high negative emotionality and high activity) was significantly associated with only lower ML grades, with a difference of 1 *SD* in the reactivity score being now associated with a difference of -0.09 in ML grades.

⁷ The results of previous analyses examined separately for teacher- and self-rated temperament are presented in original article (Study IV).

Table 10B. Associations of teacher-rated and self-rated teachability traits (concurrently, one trait at a time) with to Mathematics and Mother language grades. Separate random-intercept multilevel linear regression models.

Teachability factor adjusted for student and teacher gender and teacher age	Teacher-rated			Self-rated		
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>
Mathematics						
Task Orientation	0.58	0.04	<.001	0.13	0.03	<.001
Reactivity	-0.54	0.04	<.001	0.00	0.04	0.96
Personal-Social Flexibility	0.24	0.05	<.001	-0.07	0.04	0.13
Mother language						
Task Orientation	0.51	0.04	<.001	0.14	0.04	<.001
Reactivity	-0.32	0.04	<.001	-0.09	0.04	0.04
Personal-Social Flexibility	0.20	0.05	<.001	-0.04	0.05	0.37

Note. *Ns* for Math = 43 teachers and 636 students. *Ns* for Mother language = 26 teachers and 427 students.

β = Standardized regression coefficient; All results for teachability and school grades are presented for standardized scales (Mean = 0, *SD* = 1), adjusted for teacher and student gender and teacher age.

Each model includes teacher-rated and self-rated teachability factor.

Covariance type VC (Variance Components) was chosen for all analyses.

Intercept parameters (all *ps* < .001) and covariance parameters (all *ps* < .05) are omitted from the table.

4.3.3 Summary of the results of Study IV

To summarize, the results are as follows: (a) teacher-rated temperament and teachability were more strongly associated with students' Math and ML grades than self-rated temperament and teachability; (b) all the teacher-rated temperament traits and teachability components were associated with school grades, accounting for 38% and 45% of the variance in Math and ML, respectively; and (c) teacher perceptions of student temperament did not show as fine-grained a structure as self-rated temperament.

Of the self-rated temperaments (d) high task orientation and its components, high persistence and low distractibility, were associated with better school grades both in Math and ML; (e) low reactivity and low negative emotionality were associated only with better ML grades, whereas high inhibition was associated only with better Math grades; and (f) students' self-ratings produced a more differentiated factor structure than teacher-ratings.

5 Discussion

5.1 Are there gender differences in teachers' perceptions of student temperament, educational competence, and teachability?

The major finding was that teachers (both females and males) perceived boys' temperament, EC, and teachability to be less appropriate to the school context than girls'. In general, teachers' perceived boys' activity, negative emotionality, inhibition, and distractibility as significantly higher, but persistence, mood, and EC as significantly lower, than those of girls. This is in line with previous studies (Keogh, 1982; Martin et al., 1994; Mendez et al., 2002; Sanson et al., 1994; Walker, Berthelsen, & Irving, 2001) and confirms our hypothesis. Moreover, it confirms the lower level of teacher-perceived boys' 'task orientation' involving high activity, high distractibility, low persistence, and low mood, referring also to boys' lower teacher-perceived teachability (Keogh, 1982; Keogh et al., 1982). This temperament pattern is also consistent with a 'difficult' temperament as defined by Thomas and Chess (Thomas & Chess, 1977). Teachers usually perceive 'difficult' students, mostly boys, as irritable and annoying, because they may have problems in adapting to classroom routines and changes (Keogh, 2003; Thomas & Chess, 1977). This is seen as indicating problems in boys' compliance with teacher demands (Stuhlman & Pianta, 2002) and as causing more conflictual relationships with their teachers (Hamre & Pianta, 2001; Rudasill & Rimm-Kaufman, 2009; Saft & Pianta, 2001). Furthermore, this temperament pattern makes boys generally more vulnerable to a negative teacher–student relationship (Rudasill & Rimm-Kaufman, 2009; Silver, Measelle, Armstrong, & Essex, 2005), which in turn has been shown to lead to lower school grades (e.g., DiLalla et al., 2004; Hamre & Pianta, 2001), especially if the teacher and the student are of opposite gender (Dee, 2007). The results support the possibility of these unwanted consequences, although the association between teacher-perceived student temperament and the teacher–student relationship was not investigated.

In addition, the results replicated previous findings which showed that teacher-perceived students' low task orientation is associated with teacher-perceived students' low EC (Keogh, 1989, 1994). Here it was true particularly for boys. The present results suggest that teachers might include student EC in their evaluations of temperamental task orientation (Keogh, 1989, 1994). However, boys with low task orientation may be similar in cognitive ability, as well motivated, and as mature as girls, but due to their tempera-

ment differ only in how they respond to the demands of the school (Keogh, 2003). It is also possible that task orientation, particularly activity and persistence, may actually be associated with students' higher motivation. Although the present results cannot give affirmative evidence of teachers' underestimating the EC of boys with low task orientation, it is obvious that these boys do not match teachers' ideas of what an appropriately behaved student should be. Thus the results raise doubts about the position of boys' and their 'goodness of fit' climate (Chess & Thomas, 1999; Keogh, 1986; Pullis, 1989) in the school environment.

5.1.1 Does teacher gender matter?

Three major findings were as follows: (a) although teachers perceived boys' temperament and EC to be less appropriate in the school context than girls', the differences between boys and girls were not as large when perceived by male teachers as they were when perceived by female teachers; (b) males perceived boys' temperament, EC and teachability to be more appropriate in the school context than females; and (c) males were also stricter regarding their perceptions of girls' traits.

The findings highlight the role of male teachers and their pedagogy, when considering the school life of both genders. Although boys were perceived as lower in persistence and EC and higher in distractibility and inhibition than girls by teachers of both genders, the gender gap narrowed between boys and girls when a boy student was rated by a male teacher. This means that male teachers perceived girls' and boys' persistence, EC, distractibility, and inhibition to be closer each other than female teachers. Furthermore, male teachers perceived boys' temperament to be more appropriate for school than female teachers and also viewed them as more capable in EC and teachability. In addition, male teachers' ratings of girls' activity and negative emotionality were significantly higher, and persistence and EC significantly lower, in comparison with female teachers' ratings for girls.

The results support Bettinger and Long's (Bettinger & Long, 2005) and Dee's (Dee, 2005, 2007) findings, where same-gender matching improved except for teacher's perceptions of student capacity and performance, as well as student achievement and engagement with the teacher's subject. The present findings also reveal a positive role model effect (Bettinger & Long, 2005; Cushman, 2008; Dee, 2007; Francis, 2008), or situation where teachers show a greater capacity to understand the characteristics of students of the same gender. However, the results are contrary to those of other studies where same-gender matching has not been associated with student outcomes (Driessen, 2007; Ehrenberg et al., 1995; Lahelma, 2000), or where boys'

school attitudes have been found to be more positive when taught by female teachers (Carrington et al., 2008; Sokal et al., 2007). Consequently, it might be asked whether male teachers are also prone to practicing a more understanding and gentle pedagogy in the classroom for boys compared to girls, or in comparison with female teachers' pedagogy with boy students.

In contrast, when compared to female teachers, male teachers are likely to be stricter and more critical with regard to their perceptions of girls' traits. They may also underestimate girls' persistence and EC, or female teachers may overestimate girls' persistence and EC as well. This is contrary to the hypothesis in which it was expected that male teachers practice a more gentle and relaxing pedagogy with their students, whether boys or girls. However, this seems to be true only for boys. On the other hand, the results are in line with Dee's (Dee, 2005) findings, where female students in particular were seen as more inattentive by male teachers, although both female and male students were seen as disruptive by teacher of the other gender.

5.1.2 Does teacher age matter?

With respect to teacher age, the two major findings were as follows: (a) when a boy's teacher was an older man, teacher ratings of inhibition were higher, and ratings of mood lower, compared to a female teacher's ratings for boys or girls, but (b) the older the teacher in general, the more mature he/she perceived a student to be. The results considering the influence of teachers' age are also contrary to the hypothesis. It was expected that in general, teachers' ageing would increase their negative views of students' temperament, especially task orientation, and decrease their perception of students' EC and teachability, independent of the teacher's gender. In addition to cultural gender stereotypes and expectations (Kerr, 2001; Zhou et al., 2009), the results may be revealing in terms of the professional development and lifespan of a teacher.

The fact that teachers' ageing increases their perceptions of students' maturity may suggest that a teacher's professional and general life experience increases his/her general confidence in a student, and probably also allows him/her to be surer about a student's ability. However, male teachers' ageing, in particular, decreased their perception of boys' mood and increased their perception of boys' inhibition. It seems that with ageing, male teachers become more strict and intolerant with boys, but with female teachers, such an age-related development does not occur. This might have a significant influence also on the teacher-student relationship and interaction with boys (Rudasill & Rimm-Kaufman, 2009; Silver et al., 2005). However, because of the cross-sectional study design, it cannot be said whether teachers' perceptions

of students' temperament change with ageing, or is it a question of generational differences between younger and older teachers.

Previously, boys have been found to be more pressured than girls to change their inhibited, shy behaviour (Kagan, Reznick, & Snidman, 1987; Kerr, Lambert, Stattin, & Klackenberg-Larsson, 1994), which appears to have been the case for a long period of time in the Finnish cultural and educational climate. Thus another possible mechanism explaining the present findings is that older male teachers might carry stronger cultural expectations that reflect more traditional stereotypes of male gender.

5.2 Are teacher-perceived temperament, educational competence, and teachability associated with student school achievement?

Teacher-perceived temperament, EC, and teachability were significantly associated with students' teacher-assigned Math and ML grades, which confirm the hypothesis and is in line with the previous studies (Guerin et al., 2003; Martin & Holbrook, 1985; Martin et al., 1994). The associations were found among all temperament traits (i.e., activity, persistence, inhibition, negative emotionality, mood, and distractibility), teachability factors (i.e., task orientation, personal-social flexibility, and reactivity) and traits related to teacher-perceived EC (i.e., cognitive ability, motivation, and maturity). The associations were found with both Math and ML grades, being more apparent in Math after controlling for student and teacher gender, teacher age and student self-rated temperament.

As well, gender differences and some interactions between student and teacher characteristics in relation to school grades were noted when teacher-perceived temperament and EC were taken into account. These findings are discussed later in Section 5.2.2, *Do teacher and student gender and teacher age matter?*

Activity, persistence, and distractibility (i.e., traits that form the task orientation factor) as well as traits related to EC were associated with school grades more clearly than other traits. High distractibility was the strongest factor for lower Math and ML grades, with one standard deviation of temperament difference being associated with a -0.58 and -0.50 standard deviation difference in Math and ML grades, respectively, corresponding to about a one-half grade difference in the subjects. Likewise, high persistence was the most supportive trait for higher Math and ML grades, with one standard deviation of temperament difference being associated with a 0.62 and 0.47 standard deviation difference in Math and ML grades, respectively (nearly a half a grade difference in the subjects). Regarding temperamental activity, the association was negative for both grades, with one standard deviation of

difference being associated with a -0.45 and -0.32 standard deviation difference in Math and ML grades, respectively, whereas inhibition was shown to have the weakest influence on both subject grades. The effect sizes are in line with the literature (for a review, see Guerin et al., 1994, 2003; Martin, 1989a; Strelau, 1998) and the given boundary values (Cohen, 1988), and thus suggest moderate to rather strong associations between student temperament and school grades.

The results are in line with research which has shown that students with low temperamental task orientation, low EC, low personal-social flexibility (i.e., approach, positive mood, and adaptability), and high reactivity (i.e., negative mood, intensity of response, and reactivity) have been perceived as less capable and less teachable by their teachers (Keogh, 1982; Lerner et al., 1985; Martin & Holbrook, 1985) and have received lower school grades (Keogh, 1994; Martin, 1989a; Martin & Holbrook, 1985; Martin et al., 1994). The significant contributions of negative emotionality and inhibition as shown here have not been found previously, at least not at this intensity. These results are not, however, surprising in Finnish culture where restrained behaviour is a common and important aspect in all individuals' activities. Thus the spontaneous expression of feelings, especially negative ones, is somewhat inadvisable in the Finnish cultural and educational climate.

When interpreting the findings it should be noted, however, that at least five other mechanisms might explain the associations between temperament and school achievement. First, temperament has been found to be associated with students' working style (Guerin et al., 1994; Kristal, 2005; Rothbart & Jones, 1998), selected problem-solving strategies (Davis & Carr, 2002), and willingness to approach certain learning tasks (Caspi, 1998; Caspi & Shiner, 2006; Keogh, 2003; Kristal, 2005; Martin, 1989a). Therefore certain temperament traits, such as traits related to task orientation, approach, and quality of mood, might either support or hinder students' successful completion of given learning tasks and overall progress in learning.

Second, student temperament may be associated with students' ability to adapt to the many expectations and demands (for example, sitting still and concentrating on tasks for long periods) set by the school environment.

Third, the prevailing national assessment instructions might encourage teachers to include innate temperament as a part of evaluation criteria for students' school grades. This is particularly likely considering the assessment of a student's working style, learning process and class activity, which must be evaluated according to national guidelines, but which is, however, mainly based on innate temperamental behaviour styles.

Fourth, teacher-perceived temperament might bias teachers' assessment of student school achievement (the so-called 'halo effect'). It would thus be important to reconsider the prevailing national evaluation criteria and practice, and increase teachers' professional knowledge of innate temperament and its association with student behaviour and working styles.

Fifth, it is also possible that the 'halo effect' runs in the opposite direction, meaning that temperaments of talented or high-ability students are evaluated more positively and vice versa. This is possible because of the cross-sectional study design and because we did not adjust for students' standardized cognitive ability test performances during the analyses. Recently in Finland, however, Hintsanen and colleagues (Hintsanen et al., 2012) found an association between teacher-perceived temperament and students' school grades even when students' cognitive abilities were taken into account. They therefore suggested that this kind of 'halo effect' would be improbable.

5.2.1 What is the role of educational competence?

High teacher-perceived EC (i.e., cognitive ability, motivation, and maturity) was associated with high Math and ML grades. Moreover, teacher-rated EC mediated the statistically significant association of the above-mentioned six temperament traits with Math and ML grades. As in previous studies (Keogh, 1983, 1994), this might suggest a 'halo effect'; that is, teachers are likely to interpret these temperamental factors as reflecting maturity and motivation. This applied to both Math and ML. However, it is also possible that some temperament traits, such as activity, persistence, and positive mood may be associated with higher student motivation, which in turn may increase student task-oriented behaviour. Consequently, this may appear in the view of teachers as a student's more mature and capable behaviour. Recent research has also shown a rather low but significant ($r = .18, p < .01$) correlation between student self-rated temperamental persistence and standardized cognitive ability test score (Hintsanen et al., 2012), which indicate possible associations between students' objective cognitive ability and task-oriented behaviour. However, this still remains speculative because it is also possible that students with high persistence perform better in standardized cognitive ability tests because they can apply themselves longer and harder towards accomplishing given tasks.

Further, activity interacted significantly with the level of EC. Among students with high levels of EC, activity was negatively related to Math grades, whereas activity was unrelated to Math grades among low EC individuals. Despite the significant EC \times persistence and EC \times distractibility interactions,

the associations of persistence and distractibility with the Math grades were similar on both levels of EC.

The results also reveal the complex role of activity. Among students with high teacher-perceived EC, high activity seemed to be an obstacle to achieving high grades while this was not true in students with low EC. It might be that teachers perceive students' temperamental activity differently in different contexts; however, it seems to be negative in the context of high EC. A possible explanation might be that high activity level has often been found to be associated with negative emotionality and low persistence (Martin & Bridger, 1999). It is also possible that if the teacher perceives a certain student as low in cognitive ability as well as low in motivation and maturity, the role of the student's activity in the teacher's perceptions may be less important in general.

5.2.2 Do teacher and student gender, and teacher age matter?

There were no gender differences in students' Math grades, and boys' ML grades were almost one grade lower than girls'. However, when temperament and EC traits were adjusted for, boys received lower ML and higher Math grades in comparison with girls. Teacher gender had no association with students' school grades and teacher age was associated with student school grades only in ML.

Teachers rated students' ML grades significantly lower in boys, and almost two-thirds (62%) of this difference could be explained by gender differences in teacher-perceived temperament and EC. Gender differences in Math, by contrast, emerged only when gender differences in teacher-perceived EC and certain temperament traits (e.g., persistence and distractibility) were taken into account. However, when girls and boys with similar teacher-perceived EC and temperament characteristics were compared, boys achieved higher Math grades.

Considering ML, this is in line with the original hypothesis and with previous studies confirming boys' lower teacher-assigned grades in comparison with those of girls in stereotypically feminine subject area, such as reading, spelling, and writing (e.g., Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Pomerantz, Altermatt, & Saxon, 2002). Teachers may perceive girls' temperament to be more appropriate for ML instruction and unconsciously carry this perception over to giving better ML grades. It is also possible that girls' temperament profiles might actually assist girls toward better achievement, because girls' temperaments may help them to better adjust to the many demands of the school environment than boys' temperament. Better ability to concentrate in a noisy environment, for example, helps student's

general performance and progress in school tasks, which may again increase girls' possibilities to become more motivated in the given tasks.

The results for Math imply that temperament traits may suppress some gender differences in school achievement that might surface if boys and girls had similar temperament characteristics. Thus the results support the current concern over male students' learning circumstances, the 'goodness of fit-climate' (Keogh, 1986; Pullis, 1989; Thomas & Chess, 1977), and general well-being in the school environment.

In ML it was not possible to demonstrate the role of teacher gender owing to the lack of male teachers, which also reveals the numerical imbalance between female and male teachers in Finland. However, in ML, female teachers assigned more importance to positive mood in boys than in girls. The results considering teachers' age suggested that younger teachers might assign more importance than older teachers to student temperament and teacher-perceived EC when grading students'.

Student negative emotionality, EC and motivation were less strongly related to ML grades in older compared to younger teachers. This was also the case for maturity and inhibition in boys but not in girls. This finding might parallel results presented by Van Houtte (Van Houtte, 2007), who reported that older teachers showed more trust in their pupils independently of the pupils' teacher-perceived teachability. Perhaps older teachers are better at separating temperament from their assessments of students so that temperament assumes a lesser role than in assessments made by younger teachers. It should, however, be noted that the cross-sectional study design, which prevents examining temporal relations, permits only the reflections mentioned above, not affirmative evidence for the given results.

5.3 Which is the stronger factor in student school achievement—teacher-rated or self-rated student temperament and teachability?

When examined with both the construct of teachability and with the six different temperament traits, the main findings can be summarized as follows: (a) compared to the students' self-rated temperament the associations between teacher-rated temperament and school grades were stronger; (b) the teachers rated temperament so that it did not show as a fine-grained structure as compared to the students' self-ratings, which were more differentiated, i.e., divided into more distinctive temperament factors reflecting the three factors of teachability; (c) all teacher-rated temperament traits were associated with school grades in both Math and ML; while (d) of self-rated traits only task orientation and its components, persistence and distractibility, were associated with both Math and ML grades, whereas reactivity and negative emo-

tionality were associated only with ML grades and inhibition only with Math grades; and (d) in teacher-ratings, the overlapping between temperament and school grades was strong, whereas student-rated temperament was less strongly associated with school grades.

It was observed that all the associations between teacher-rated temperament and school grades remained significant and were not essentially attenuated even after controlling for the students' self-rated temperament, suggesting that teachers' perceptions of student temperament are associated with students' school performance independently of students' self-rated temperament. This was true after analyses with both the three factors of teachability and six different temperament traits. The teacher-rated task orientation component and traits related to the ideal 'school temperament', that is, high persistence, low activity, and low distractibility, were the most pervasive positive factors for better Math and ML grades, although these associations were somewhat stronger for Math.

In proportion, teacher-rated reactivity along with high negative emotionality and high activity were the most pervasive negative factors for lower Math and ML grades, the associations being again stronger for Math. Furthermore, as a component of teachability, teacher-rated personal-social flexibility was the less pervasive factor in relation to student school grades in both subjects. In addition to the present results being in line with our hypotheses, they confirm previous findings pointing to a rather strong association between teacher-rated temperament and school outcomes (for reviews, see Guerin et al., 2003; Hintsanen et al., 2012; Keogh, 2003; Kristal, 2005; Martin et al., 1994; Zhou et al., 2010), and the very influential role of task orientation in this relationship (Guerin et al., 2003; Keogh, 1983, 1994; Martin, 1989a; Martin et al., 1994; Martin, Drew, Gaddis, & Moseley, 1988). However, previous studies have not been able to compare the relative strengths of teacher- and self-rated temperaments analysed in the same model, which may be seen as an advantage of the current study.

The results also suggest that teachers perceive student temperament as being a less fine-grained structure compared to the students' self-ratings, so that different temperament traits are not as clearly differentiated. Teachers' perceptions of student temperaments are likely to be more context-specific than students' self-ratings, as teachers' perceptions tap primarily into the students' classroom behaviour in school (Keogh, 2003, 1994). In addition, because of the close connections with working styles, the traits of task orientation (persistence and distractibility) and reactivity (activity and negative emotionality) may be more pervasive in the school than home context. Students high in task orientation and low in reactivity can maintain their attention in given tasks,

finish their tasks despite interruptions, and adapt to repeatedly varying learning situations (Keogh, 1983). This means that their temperament allows them to moderate and direct their school performance according to the given tasks (Keogh, 1983).

However, in addition to student's high activity and distractibility being related to more negative attitudes of the teachers (Martin et al., 1983), the present study showed that high activity and high negative emotionality is associated with teachers' more negative perceptions of student's EC (i.e., cognitive ability, motivation and maturity). This indicates that teacher-rated temperament may reflect a specific assessment of student temperament which may differ in its predictive validity as compared to the temperament ratings made by others or in different contexts.

Moreover, in the teacher-ratings the overlapping between temperament and school grades was stronger than in the student ratings, which might reflect a 'halo effect'. This is in line with previous findings in which teacher reports have been found to be affected by the teachers' impressions of student classroom behaviour and achievement (Keogh, 1982; Martin et al., 1994). In the Finnish educational system this is especially possible because the national evaluation guidelines (FNBE, 2004) suggest that teachers assess the student's classroom behaviour, working styles and learning process in different areas of the subject and thus enable a single teacher's subjective assessment to have a large influence. However, this may complicate student school evaluation, which should be based solely on a student's actual achievement, independent of the student's temperament and personality. This pathway between student temperament and school outcomes has previously been illustrated by Martin and his extensive studies (Martin, 1992). In addition, Caspi and colleagues (Caspi, 1998; Caspi, Elder, & Bem, 1987, 1988; Caspi & Silva, 1995) have presented the concept of 'cumulative continuity', where temperament-related consequences may be combined and elaborated over the course of time.

The students' self-ratings produced a more differentiated factor structure of temperament than the teacher-ratings, and the associations between the student-rated temperament traits and school grades were more specific and in line with the subject-related tasks and working habits as expected based on the literature (Guerin et al., 2003; Keogh, 1982; Kristal, 2005). Of the self-rated temperament factors, high task orientation, and high persistence and low distractibility as single traits were significantly associated with both better Math and ML grades. On the other hand, a high self-rated reactivity factor and high negative emotionality as a single trait were significant only in relation to lower ML grades, whereas high inhibition was significant only in

relation to better Math grades. This was true even after adjusting for the influence of teacher-rated temperament, and for teachers' and students' gender and teachers' age. However, as a wider component of teachability, self-rated personal-social flexibility was not shown to be significant.

Overall, the self-rated associations were small, with magnitudes varying from low to moderate, but these were in line with the literature (Guerin et al., 2003; Keogh, 1982; Kristal, 2005). This further confirms the important role of task orientation for both school subjects independent of the rater (Guerin et al., 2003; Keogh, 1989; Martin, 1989a) and the association of self-rated inhibition with the Math grades (Hintsanen et al., 2012). Students' temperamental inhibition has previously been found to affect Math achievement through working styles and retrieval strategies that students are prone to choose in approaching problem-solving tasks (Davis & Carr, 2002). The significant role of self-rated reactivity and particularly the role of negative emotionality in ML grades have not been demonstrated earlier, at least not on this scale. Because Finnish culture encourages humble and composed behaviour and the controlling of feelings, negative emotionality may be seen as inappropriate behaviour in relation to successful school achievement particularly in such a subject as in ML where instructions contain many socially active cooperative forms, like speaking aloud, creative acting and asserting one's own thoughts in general.

There are at least three other possible explanations for the results of self-rated high inhibition being associated with better Math grades and that reactivity and high negative emotionality was associated with lower ML grades. First, the Behavioural Inhibition System (BIS) has been found to be positively associated with the Deliberation subscale (i.e. tendency to think matters through before acting or speaking) from the Conscientiousness dimension of the Five-Factor Model (Keiser & Ross, 2011). This may indicate that a student with high inhibition may not rush through Math problems or record the first possible alternative solution, but instead work more deliberately and thus benefit from his or her higher inhibition.

Second, Math classes may be more formal and more related to quiet task-centred activities than ML classes. In contrast, ML classes may incorporate more free-form working methods, even drama and creative articulation and other social and cooperative activities, which are not expected as much in Math instruction. Therefore a subject-related working culture may be favourable for students with higher inhibition in Math and lower negative emotionality during ML instruction.

Third, reactivity and negative emotionality may be more visible during the socially active situations of ML instruction than in the more structured

situations of Math, where single tasks may take longer to complete. However, in the current study, it was surprising that self-rated high inhibition was not associated with lower ML grades. Likewise, it is important to note that considering negative emotionality in ML, the bivariate correlation between self-rated and teacher-rated negative emotionality was very low and not significant ($r = .04$), which raises the question of what teachers, in proportion, have intended in terms of negative emotionality as compared to their students' intentions.

Although the bivariate correlations between teacher- and self-rated temperaments varied only from modest to moderate (details of the results are presented in the original article), they proved to fit at the same magnitude with other cross-informant data (for example, youth-parent-, parent-parent-, parent-teacher-, and teacher-self-ratings) previously found in several areas of study and wide meta-analyses (Achenbach et al., 1987; Achenbach, Dumenci, & Rescorla, 2002; Achenbach et al., 2005; Vazire & Mehl, 2008).

The modest student-teacher agreement in the current study implies, however, that the teacher- and self-rated perceptions of student temperament are not interchangeable. They may reflect different roles, viewpoints, and social contexts of the rater; for example, because of daily repeated contacts with students, teacher-ratings may reflect teacher's opinions of student school performance as regards working styles whereas it has been suggested that students' self-ratings may be more affected by views held by their peers (Hintsanen et al., 2012). This is in line with the results of wide meta-analyses conducted by Achenbach and his colleagues (Achenbach et al., 2005), where they argued that each rating is prone to be affected by the situation- and framework-related factors, the attributes of the rater, and the nature of the relationship between the informant and the student. Thus the low correlations indicate that it is important to use more than one informant and to take into account who has been the informant when interpreting the findings.

On the contrary, there are also other possible explanations for the dissonance between student-teacher agreements. It is always possible that a teacher might not know enough of his/her students' personal characteristics, especially if the teacher is new, has several teaching groups, and meets students only once or twice a week in restricted circumstances. However, in the current study this kind of explanation is unlikely because teachers' ratings were implemented in compliance with the condition that the student had been attending the same class taught by the same teacher for more than one year and knew the student well. In addition, it should be noted that students' young age and strong developmental stage might influence or cause variability in their self-perceptions.

Regardless of this dissonance between student-teacher agreements, the importance of a teacher's perceptions and attitude vis-à-vis a student cannot be either diminished or denied. In addition to the findings that previously a student's temperament has been shown to influence how the teacher tolerates the student (for reviews, see Guerin et al., 2003; Kristal, 2005; Martin, 1989b), the present study indicates that there is also an association between student temperament and teacher-assigned school grades. It is possible that a certain temperament may lead to a teacher harbouring negative attitudes toward that student, in which case the student may receive a negatively biased treatment from his/her teacher. This may influence a student's general and academic self-conception as well as how he/she experiences and orients towards others (Guerin et al., 2003; Harter, 1986, Harter, 1999; Klein, 1995; Kristal, 2005; Martin, 1992, 1994). However, as mentioned earlier (for example in Section 5.2 *Are teacher-perceived temperament, educational competence, and teachability associated with student's school achievement?*), there may be many other avenues and explanations in the association between student temperament and teacher-assigned school achievement. For more detailed and reliable implications, longitudinal studies with multiple raters are needed to examine the temporal relations of the associations found in this study.

5.4 Are there subject-related differences in the association between teacher- and self-rated temperament and teachability with student school achievement?

The associations of teacher- and self-rated temperament with school grades were stronger throughout in Math than in ML. This is in line with our hypotheses and with the findings of Maziade and colleagues (Maziade et al., 1986) but not with the majority of the other findings, which have reported stronger associations in ML (Guerin et al., 1994; Martin, 1989a; Martin & Holbrook, 1985; Newman et al., 1998; Strelau, 1998). Because of its versatile domains (e.g. drama, creative writing, literature, and communicative skills), the subject-related pedagogy of ML in Finland may not require students to adopt as exact working habits as is expected in Math. It was therefore suggested that when subject-related environmental requirements are stricter and more limited in as Math, the importance of the compatibility (i.e., goodness of fit) or dissonance (i.e., poorness of fit) (Chess & Thomas, 1999) between the learning circumstances of the school environment as well as the student's own behavioural characteristics, working style, and capacity is strengthened. Consequently, the role of temperament emerges as more visible in Math than in ML. It is also possible that over the course of time, the working methods in

ML instruction have changed from a strict emphasis on grammar requiring rote memory tasks to more creative working styles. Therefore in present-day studies, the previously reported associations between temperament and ML may no longer be as strong as earlier supposed. However, the current results expand previous findings by offering evidence of the broader influence of teacher-perceived temperament and teachability on students' Math and ML grades as examined at the same time.

5.5 Methodological considerations

The main strengths of the study are a nationally representative sample of Finnish students and a very low attrition rate. As a rule, population-based studies where students' assessment has been considered from perspectives other than cognitive ones are unusual. Finnish comprehensive school is the most appropriate "real-life laboratory". The whole age group can be contacted because there are no private schools or parallel school systems. Teachers are similarly educated and all schools follow the same national curriculum, which offers optimal circumstances for investigating individual differences and variances in the school context. On the other hand, the homogeneity of the school system enables a low variance in terms of school and teacher education systems. Thus the results are internationally comparable with findings from the other OECD countries.

Another strength is that the current study seems to be the first to examine associations between teacher- and self-rated temperament and teachability and both Math and ML grades at the same time, as a function of student and teacher gender and teacher age, and in the same population-based sample. Thus the study contributes important new findings and expands the literature.

The third strength is that the use of more than one informant reduces the conceivable problems produced by the common method variance (Kline, Sulsky, & Rever-Moriyama, 2000).

The main limitation is that the study has a cross-sectional design, which does not allow for conclusions about the temporal directions or the causality of the associations. Therefore it is also possible that students' school grades might be associated with teachers' perceptions of student temperament and EC. There may also exist some other latent variables, which may be associated with both temperament and school grades, and explain the link between temperament and school grades. Negative life events, for example, could function as such third factors. Further, multicollinearity between some variables (e.g., between the traits of task orientation) weakens the determination and accuracy of the results. However, this cannot be seen merely as a limitation of the study design because the main purpose was to investigate how

these variables are associated with teachers' perceptions in practical school life in spite of causality.

Second, the research focuses only on one age cohort. Due to the strong developmental and school stage transition phases, additional research is needed to generalize these results for the different age groups from the 15-year-olds included in the study.

Third, it was not possible to verify the students' self-reported grades against official school registers, which would be advisable because self-reported grades may also contain a certain bias (Kuncel, Credé, & Thomas, 2005). However, students tend to report their academic performance quite accurately when they assume that this will be checked independently (Gray & Watson, 2002). The students in the current study knew that one to three of their subject teachers would fill a teacher-reported form on their temperament and school performance.

Fourth, because of the gender difference in teacher proportion, it was not possible to demonstrate the distinct association between teacher gender and ML, which partly limits the conclusions regarding ML.

Fifth, the study examined teacher age and gender only as potential modifying factors in the association between teacher-perceived student temperament and school achievement. Other teacher characteristics, such as their personal beliefs about teaching, their personality and character, and their educational background, might be associated with the relationship between temperament, EC, and school grades, and thus may also be relevant here. Such characteristics need to be considered in future studies on the topic.

Finally, it was not possible to take the role of parental socio-economic status into account, which is an important limitation. Maternal education in particular might explain a significant portion of students' school achievement, and function as a confounder between temperament and school grades. In addition to taking into account the above-mentioned six factors, longitudinal research designs are needed to investigate the cross-time linkage between temperament and academic achievement assessed by different raters.

6 Conclusions and Practical Implications

The main conclusions are as follows: (1) teacher-perceived student innate temperament, EC, and teachability explain the rather high proportion of teacher-assigned students' school grades; (2) the teachers' ratings seem to carry a systematic variance both by their own gender and the student gender, and by their own age; (3) the variance between teacher-rated and self-rated temperament might be explained by teacher-ratings and self-ratings measuring different concepts and explaining different variance, and (4) because of these three aspects, it is possible that the grade giving might be biased by teachers' perceptions of student innate temperament. However, the study cannot give affirmative evidence for this mainly because of it being cross-sectional research, which prevents examining the temporal directions and causality of the associations, and because we did not directly examine how teachers' opinions of students' temperaments affected their grade giving.

The present findings suggest that teacher- and self-reports are not merely equal and alternative methods of assessing student temperament, but they might reflect different concepts connected to different social contexts. Both teacher- and self-ratings can be used, but they measure and explain different variance. Consequently, which ratings are the more practicable or accurate remains a matter of speculation. This is why it is important to obtain multi-informant data on student characteristics and to consider both the rater and the purpose of the rating when evaluating research findings on student temperament. In future studies the associations between teacher- and self-rated temperament traits and objective test results should be compared with the comparative associations with the grades given by teachers in order to have more information on the possible bias in grade giving. This topic in grade giving is an important question to examine in future studies, but reliable evidence can only be gathered by longitudinal research with multiple raters.

The findings may point to some individual and societal consequences and suggest some practical implications considering future teacher training and school education.

6.1 Individual and societal consequences of the current findings

6.1.1 Boys' exclusion

The findings share the international concern about boys' affirmative goodness of fit-environment in school, and raise the question of whether the prevailing education system is supportive and equitable for both genders (Chess

& Thomas, 1999; Keogh, 2003; Kristal, 2005; Thomas & Chess, 1977). In spite of the excellent PISA results (OECD, 2004, 2007a, 2007b, 2010), almost nine percent of Finnish students, mostly boys, drop out of studies after ninth grade (Tilastokeskus, 2011). Thus boys' general school well-being is an important issue both domestically and internationally.

The dissonance (i.e., poorness of fit) between student behavioural characteristics and learning circumstances is particularly critical with respect to boys' eligibility for future studies. With all its societal and health implications, the most far-reaching consequence of 'poorness of fit' climate is boys' risk of dropping out of society (Martin, 1992). It seems that the general school expectations might not be appropriately matched with boys' behavioural characteristics and needs.

6.1.2 Girls' later underachievement

Another socially important issue is girls' underachievement in later professional education and careers despite their excellent school success and advantage in comparison with boys. Although girls appear to meet the expectations and demands set by the school environment better than boys, the school reports' promises of girls' educational predominance and more preferable careers fail to materialize later on. Instead, it seems that girls' capacities may become sidelined. This can be explained by several reasons, for example by motivation, attitudes, and the values reflected by family and society. As such, the present school system may not sufficiently acknowledge the best characteristics of both genders in the successful choice of a later career.

6.1.3 Fairness of assessment

The present results suggest that more consideration be given to the prevailing assessment practice in the Finnish education system where the student's learning process with lesson activity occupies a central position. The temperament-based approach suggests that there is no single proper way to study, learn, and complete learning tasks and processes. The possibility to follow one's own temperament and behavioural style in studying is seen as the most effective way for every student (Keogh, 2003; Kristal, 2005; Thomas & Chess, 1977), while strong demands and strict expectations related to learning style and process might be even detrimental.

Teachers' verbal assessments in school reports of students' learning processes are especially problematic, because students' temperaments are involved in these evaluations. Verbal assessments focus on the working styles, but they refer to temperamental characteristics, such as, "you are active and

eager”, “you work in a motivated and target-oriented way”, “you are good at concentrating for a long time on a given task”, “you are persistent in your tasks”, “you come to school in a good mood”, “you are socially active in group tasks”, etc. Although it may be important particularly for the younger students to receive feedback related their working styles, this kind of feedback mostly concerns individual personality, which may, particularly if it is negative, disturb a student’s developing self-image and even obscure the student’s ability to recognize his/her actual cognitive strengths. It is especially problematic if temperament-related assessment remains a part of a teacher’s evaluation practice through the student’s later school years, affecting also teacher-assigned school grades. This might have far-reaching consequences for the student’s later education and career.

Although students’ education and teaching for certain learning processes are essential functions of the school, it would be advisable, however, that these teaching areas be kept apart from subject assessment, school grades, and GPAs. In addition, a student’s lesson activity does not absolutely measure the student’s academic know-how, but for the most part only his or her innate temperament. Thus if it is seen as necessary to evaluate students’ learning processes and lesson activity, students could receive two separate subject-dependent school grades, i.e., for academic achievement and working style. In addition to this being informative, it would also be an equitable assessment for students with diverse temperaments.

Everything that is estimated and measured in the school environment should be based on the official curriculum, defined targets and accepted evaluation criteria. On the other hand, whether it is appropriate to evaluate everything is an issue that should be resolved. The reliability and the validity of the measurement are important in all school evaluation.

6.2 Knowledge of temperament in teacher training

Finnish teacher training and the general education system are world-famous and of high quality. However, this research indicates some future challenges. First, the findings suggest that Finnish teachers need more information about the importance and influence of temperament in the school context. This is important in order to prevent a possible ‘halo effect’, and to differentiate cognitive skills and academic achievement from behavioural styles, and consequently ensure equitable treatment and assessment for all students, which is in accordance with their actual abilities but independent of their temperament and personality. In this respect, teachers and their education are key, and a template for any possible progress.

Second, temperament-conscious education should be taken into account in future teacher training. Knowledge about temperament would then become as an effective tool of teachers' personal practical theory and help them to become sensitive to a student's temperament, and then recognize, accept, and trust his/her own observations of that student's innate behavioural style. The current results imply that teachers may develop such sensitivity with age and experience. Temperament knowledge would therefore be especially valuable for younger and novice teachers at the beginning of their careers, for whom student temperament may assume a disproportionately large and salient role when assessing student achievement. Teachers who are sensitive and responsive to their students' innate behavioural style can cope more effectively with their students' temperaments by modifying the students' behaviour and learning circumstances as well as their own behaviour and teaching methods in the classroom. In addition, temperament-conscious teachers will have a greater tolerance of their students' behaviour and more sensitivity in recognizing students' gender-specific differences. Hence, temperament knowledge-based teacher training might function as a buffer against professional burnout or moving to other work, which is common in the teaching profession in Finland (Almiala, 2008).

Third, teachers would benefit if they increased their professional capital by understanding their own temperaments and how their temperament traits affect their students, teacher-student interactions, and educational decisions (Keogh, 1982, Keogh, 2003; Kristal, 2005). This may be even more demanding for teachers than teaching itself, but it is recommended that teachers would be at least aware of these mechanisms and effects regarding their own temperaments, as well as how different temperaments interact with one another in the classroom. This knowledge would help them to become more aware of the teaching and working methods that are the most natural to their own behavioural styles, and to realize how much both students and teachers can even tolerate the dramatically changing situations in classroom life (Hegvik, 1989; Keogh, 2003; Kristal, 2005; Martin, 1989b; Pullis, 1989). Such awareness and understanding would also help teachers to understand the interaction between colleagues in teachers' rooms, for example, as well as other adults in the school environment.

Fourth, it would be helpful for teachers to receive thorough education with psychometric information and ethical principles in conducting students' school evaluation in general, and particularly with regard to school tests and teacher-assigned grades. This education might be given during basic teacher training but also continue periodically during a teacher's career.

6.3 Temperament-conscious education

Although Finnish upper-comprehensive school students have continually ranked highly in international PISA assessments with regard to Math, reading, and science (OECD, 2004, 2007a, 2007b, 2010), temperament-conscious education in the school context might even increase the effectiveness as well as the enjoyment of school life (Haapasalo, Välimaa, & Kannas, 2010; OECD, 2004; Välijärvi et al., 2007) among both students and teachers.

In the view of temperament research, it should be kept in mind that similar school settings are perceived in different ways by different students (Rothbart & Jones, 1998) and therefore students' equal treatment may not be equitable treatment in the school context (Keogh, 2003; Kristal, 2005; Chess & Thomas, 1999). Temperament-conscious education encourages educators to understand and respect students' individual temperamental differences (e.g. in working styles) and to take this into account in their teaching-learning processes. This means that respecting different temperaments equally means equitable treatment for all students. However, temperament-conscious education should not be used to segregate students or for tailored classes and special teaching groups, but rather to create an educational climate with flexible learning circumstances and educational methods that fit all types of temperaments and not only for those who happen to have a so-called affirmative 'school temperament' with high task orientation (for reviews, see Hegvik, 1989; Martin, 1989b; Pullis, 1989).

Keogh (2003), Kristal (2005), and Rothbart & Jones (1998) have discussed broadly and in detail the possibilities of improving the temperamental fit in schooling. For example, creating such varying learning situations where it is possible to move and act instead of sitting still would help very active students to fit in with the rest of the class. Rothbart and Jones (1998) have stressed, that students' temperament-based differences in emotional reactivity, energy level, motivation, and attention are significantly associated with their learning and especially with their ability to demonstrate what they have learned. Therefore a deep understanding of a student's temperament and its relation to his/her classroom behaviour would be as important as knowing the student's cognitive capacity (Rothbart & Jones, 1998). Teacher-managed classroom flexibility, emotional support and acceptance may also protect students' academic performance from the harmful influence of lower temperamental attention (Rudasill et al., 2010). Consequently, this would likely improve the quality of the teacher-student interaction and relationship, which may be associated with student coping (Ruus et al., 2007). For example, students who tend to be easily overwhelmed by ambient noise usually experience more fear and discomfort, which might weaken their cognitive processes

as well as positive classroom emotions and memories (Rothbart & Jones, 1998). These students might benefit from a careful planning of the classroom structure (Keogh, 1994, 2003; Kristal, 2005), which would allow them to become familiar with and adapt to novel situations, including school tests (Rothbart & Jones, 1998). Temperament knowledge might produce the most useful results when used as a supportive and comprehensive practical tool throughout the entire teaching, studying, and learning process, and particularly teachers and students working together. Temperament-conscious teachers can help students to know their strengths better and provide tools to cope with their weaknesses. For themselves, students can improve their self-knowledge and self-control by understanding their own temperaments and the reactions that their temperaments may arouse in themselves, their classmates and their teachers in various school situations. Temperament-conscious education might then help students to succeed with respect to the most important goal of education: learning to learn.

6.4 Other remarks

The feminization of the teaching profession is a concern in a number of countries (Carrington & McPhee, 2008; Skelton, 2002) and a topical issue in Finnish education as well (Lahelma, 2000). The present research showed that female teachers' perceptions of student temperament did not differ between genders, whereas male teachers seemed to be more censorious toward girls' temperaments and more understanding of boys'. Furthermore, in terms of temperament, teacher gender did not extend its association with students' school achievement. Therefore these results at least suggest no reason to be concerned about a preponderance of female teachers in Finnish education.

Finally, temperament should not be thought of only in terms of risk issues in the school context. Temperament is not a handicap but quite the contrary: it contains many positive elements. For example, it explains why some students are cheerful, persistent, motivated, eager to learn, and willing to approach new things. Temperament-conscious teacher training and teaching practice would increase the goodness of fit of studying, future education, and later careers. As such, it would work as a supportive tool in society's attempt to prevent boys' exclusion and to improve girls' later professional outcomes so that they are in better accordance with their actual abilities and capacities.

References

- Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2002). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. *Journal of Emotional and Behavioral Disorders, 10*(4), 194–203.
- Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin, 131*(3), 361–382.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 10*(2), 213–232.
- Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the US and China: Similarities and differences. *European Journal of Personality, 7*(5), 359–377.
- Almiala, M. (2008). *Mieli paloi muualle – opettajan työuran muutos ja ammatillisen identiteetin rakentuminen*. (University of Joensuu, Publications in Education, No. 128). University of Joensuu, Finland: Faculty of Education.
- Angleitner, A., & Ostendorf, F. (1994). Temperament and the big five factors of personality. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The development structure of temperament from infancy to adulthood* (pp. 69–90). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ashley, M., & Lee, J. (2003). *Women teaching boys*. Stoke-on-Trent, UK: Trentham Books.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- Bates, J. E. (1989). Concepts and measures of temperament. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 3–26). Chichester, UK: John Wiley & Sons.
- Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review, 95*(2), 152–157.
- Boivin, M., Hymel, S., & Bukowski, W.M. (1995). The roles of social withdrawal, peer rejection, and victimization by peers in predicting loneliness and depressed mood in childhood. *Development and Psychopathology, 7*(4), 765–785.
- Boyle, H. B., & Willms, J. D. (2001). Multilevel modelling of hierarchical data in developmental studies. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(1), 141–162.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. (Advanced quantitative techniques in the social sciences series No. 1). Newbury Park: Sage Publications.
- Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York, NY: Wiley.

- Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Carrington, B., & McPhee, A. (2008). Boys' 'underachievement' and the feminization of teaching. *Journal of Education for Teaching*, 34(2), 109–120.
- Carrington, B., & Skelton, C. (2003). Re-thinking 'role models': Equal opportunities in teacher recruitment in England and Wales. *Journal of Education Policy*, 18(3), 253–265.
- Carrington, B., Tymms, P., & Merrell, C. (2008). Role models, school improvement and the 'gender gap'—Do men bring out the best in boys and women the best in girls? *British Educational Research Journal*, 34(3), 315–327.
- Caspi, A. (1998). Personality development across the life course. In W. Damon (Ed.-in-Chief) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, 5th ed., pp. 311–388). New York, NY: John Wiley & Sons.
- Caspi, A., Elder, G. H., & Bem, D. J. (1987). Moving against the world: Life-course patterns of explosive children. *Developmental Psychology*, 23(2), 308–313.
- Caspi, A., Elder, G. H., Jr., & Bem, D. J. (1988). Moving away from the world: Life-course patterns of shy children. *Developmental Psychology*, 24(6), 824–831.
- Caspi, A., & Shiner, R. L. (2006). Personality development. In W. Damon & R. M. Lerner (Eds.-in-Chief) and N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Social, emotional and personality development* (Vol. 3, 6th ed., pp. 300–365). Hoboken, NJ: John Wiley & Sons.
- Caspi, A., & Silva, P. A. (1995). Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child Development*, 66, 486–498.
- Chen, X., Rubin, K. H., & Li, B. (1995). Social and school adjustment of shy and aggressive children in China. *Development and Psychopathology*, 7(2), 337–349.
- Chen, X., Rubin, K. H., & Sun, Y. (1992). Social reputation and peer relationships in Chinese and Canadian children: A cross-cultural study. *Child Development*, 63, 1336–1343.
- Chess, S., & Thomas, A. (1999). *Goodness of fit: Clinical applications from infancy through adult life*. Philadelphia, PA: Brunner/Mazel.
- Chudgar, A., & Sankar, V. (2008). The relationship between teacher gender and student achievement: evidence from five Indian states. *Compare*, 38(5), 627–642.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coplan, R. J., Barber, A. M., & Lagacé-Séguin, D. G. (1999). The role of child temperament as a predictor of early literacy and numeracy skills in preschoolers. *Early Childhood Research Quarterly*, 14(4), 537–553.
- Cushman, P. (2008). So what exactly do you want? What principals mean when they say 'male role model'. *Gender and Education*, 20(2), 123–136.
- Davis, H., & Carr, M. (2002). Gender differences in mathematics strategy use: the influence of temperament. *Learning and Individual Differences*, 13, 83–95.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2), 158–165.

- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *The Journal of Human Resources*, 42(3), 528–554.
- DiLalla, L. F., Marcus, J. L., & Wright-Phillips, M. V. (2004). Longitudinal effects of preschool behavioral styles on early adolescent school performance. *Journal of School Psychology*, 42, 385–401.
- Driessen, G. (2007). The feminization of primary education: Effects of teachers' sex on pupil achievement, attitudes and behaviour. *Review of Education*, 53(2), 183–203.
- Drudy, S. (2008). Gender balance/gender bias: The teaching profession and the impact of feminisation. *Gender and Education*, 20(4), 309–323.
- Eaton, W. O., & Enns, L. R. (1986). Sex differences in human motor activity level. *Psychological Bulletin*, 100(1), 19–28.
- Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review*, 48(3), 547–561.
- Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, 82(5), 804–818.
- Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, 132(1), 33–72.
- Field, T., & Greenberg, R. (1982). Temperament ratings by parents and teachers of infants, toddlers, and preschool children. *Child Development*, 53, 160–163.
- FNBE. (2004). *National Curriculum for Basic Education 2004*. National core curriculum for basic education intended for pupils in compulsory education. Vammala, Finland: Opetushallitus / The Finnish National Board of Education (FNBE). Retrieved from http://www.oph.fi/english/education/overview_of_the_education_system
- Francis, B. (2008). Teaching manfully? Exploring gendered subjectivities and power via analysis of men teachers' gender performance. *Gender and Education*, 20(2), 109–122.
- Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S. et al. (1987). Roundtable: What is temperament? Four approaches. *Child Development*, 58, 505–529.
- Goldsmith, H. H., Lemery, K. S., Aksan, N., & Buss, K. A. (2000). Temperamental substrates of personality development. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 1–32). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gray, E.K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, 70(2), 177–206.
- Guerin, D. W., & Gottfried, A. W. (1994). Developmental stability and change in parent reports of temperament: A ten-year, longitudinal investigation from infancy through preadolescence. *Merrill-Palmer Quarterly*, 40(3), 334–355.
- Guerin, D. W., Gottfried, A. W., Oliver, P. H., & Thomas, C. W. (1994). Temperament and school functioning during early adolescence. *Journal of Early Adolescence*, 14(2), 200–225.

- Guerin, D. W., Gottfried, A. W., Oliver, P. H., & Thomas, C. W. (2003). *Temperament: Infancy through adolescence. The Fullerton longitudinal study*. (Longitudinal research in the social and behavioral sciences: An interdisciplinary series). New York, NY: Kluwer Academic/Plenum Publishers.
- Haapasalo, I., Välimaa, R., & Kannas, L. (2010). How comprehensive school students perceive their psychosocial school environment. *Scandinavian Journal of Educational Research, 54*(2), 133–150.
- Hall, J. A., & Halberstadt, A. G. (1986). Smiling and gazing. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 136–158). Baltimore, MD: John Hopkins University Press.
- Ham, B.-J., Lee, Y.-M., Kim, M.-K., Lee, J., Ahn, D.-S., Choi, M.-J., et al. (2006). Personality, dopamine receptor D4 exon III polymorphisms, and academic achievement in medical students. *Neuropsychobiology, 53*(4), 203–209.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*(2), 625–638.
- Hargreaves, A. (2005). Educational change takes ages: life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education, 21*, 967–983.
- Harter, S. (1986). Processes underlying the construction, maintenance, and enhancement of the self-concept in children. In J. Suis & A. Greenwald (Eds.), *Psychological perspectives on the self* (Vol. 3., pp. 137–181). Mahwah, NJ: Lawrence Erlbaum Associates.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York, NY: Guilford Press.
- Hegvik, R. L. (1989). Application of temperament theory to an individualized educational environment. In W. B. Carey & S. C. McDevitt (Eds.), *Clinical and Educational Applications of Temperament Research*, (pp. 121–124). Berwyn, PA: Swets North America.
- Hildebrandt, S. A., & Eom, M. (2011). Teacher professionalization: motivational factors and the influence of age. *Teaching and Teacher Education, 27*(2), 416–423.
- Hintsanen, M., Alatupa, S., Pullmann, H., Hirstiö-Snellman, P., & Keltikangas-Järvinen, L. (2010). Associations of self-esteem and temperament traits to self- and teacher-reported social status among classmates. *Scandinavian Journal of Psychology, 51*, 488–494.
- Hintsanen, M., Alatupa, S., Jokela, M., Lipsanen, J., Hintsala, T., & Leino, M. (2012). Associations of temperament traits and mathematics grades in adolescents: Dependent on the rater but independent of motivation and cognitive ability. *Learning and Individual Differences, 22*, 490–497.
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131–155.
- Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58*, 1459–1473.

- Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: the role of children's approach to school-work. *Developmental Psychology, 42*(1), 11–26.
- Keiser, H. N., & Ross, S. R. (2011). Carver and Whites' BIS/FFFS/BAS scales and domains and facets of the Five Factor Model of personality. *Personality and Individual Differences, 51*(1), 39–44.
- Keogh, B. K. (1982). Children's temperament and teachers' decisions. In R. Porter & G. M. Collins (Eds.), *Ciba Foundation Symposium 89, Temperamental differences in infants and young children* (pp. 269–285). London, UK: Pitman.
- Keogh, B. K. (1983). Individual differences in temperament: A contribution to the personal, social, and educational competence of learning disabled children. In J. D. McKinney & L. Feagans (Eds.), *Current topics in learning disabilities* (Vol. 1, pp. 33–55). Norwood, NJ: Ablex Publishing Corporation.
- Keogh, B. K. (1986). Temperament and schooling: Meaning of "goodness of fit"? In J. V. Lerner & R. M. Lerner (Eds.), *Temperament and social interaction in infants and children: New directions for child development* (pp. 89–108). San Francisco, CA: Jossey-Bass.
- Keogh, B. K. (1989). Applying temperament research to school. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 437–450). Chichester, UK: John Wiley & Sons, Ltd.
- Keogh, B. K. (1994). Temperament and teachers' views of teachability. In W. B. Carey & S. C. McDevitt (Eds.), *Prevention and early intervention: Individual differences as risk factors for the mental health of children* (pp. 246–254). New York, NY: Brunner/Mazel.
- Keogh, B. K. (2003). *Temperament in the classroom: Understanding individual differences*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Keogh, B. K., Pullis, M. E., & Cadwell, J. (1982). A short form of the teacher temperament questionnaire. *Journal of Educational Measurement, 19*(4), 323–329.
- Kerr, M. (2001). Culture as a context for temperament: Suggestions from the life courses of shy Swedes and Americans. In T. D. Wachs & G. A. Kohnstamm (Eds.), *Temperament in context* (pp. 139–152). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kerr, M., Lambert, W. W., & Bem, D. J. (1996). Life course sequelae of childhood shyness in Sweden: Comparison with the United States. *Developmental Psychology, 32*(6), 1100–1105.
- Kerr, M., Lambert, W. W., Stattin, H., & Klackenber-Larsson, I. (1994). Stability of inhibition in a Swedish longitudinal sample. *Child Development, 65*, 138–146.
- Klein, H. A. (1995). Self-perception in late adolescence: An interactive perspective. *Adolescence, 30*(119), 579–592.
- Kline, T. J. B., Sulsky, L. M., & Rever-Moriyama, S. D. (2000). Common method variance and specification errors: A practical approach to detection. *The Journal of Psychology, 134*(4), 401–421.
- Kohnstamm, G. A. (1989). Temperament in childhood: Cross-cultural and sex differences. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 483–508). Chichester, UK: John Wiley & Sons, Ltd.

- Krieg, J. M. (2005). Student gender and teacher gender: What is the impact on high stakes test scores? *Current Issues in Education*, 8(9). <http://cie.ed.asu.edu/volume8/number9/> [On-line].
- Kristal, J. (2005). *The temperament perspective: Working with children's behavioral styles*. Baltimore, MD: Paul H. Brookes Publishing Co.
- Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75(1), 63–82.
- Kwok, O., Hughes, J. N., & Luo, W. (2007). Role of resilient personality on lower achieving first grade students' current and future achievement. *Journal of School Psychology*, 45, 61–82.
- LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin*, 129(2), 305–334.
- Lahelma, E. (2000). Lack of male teachers: A problem for students or teachers? *Pedagogy, Culture and Society*, 8(2), 173–186.
- Lam, Y. H. R., Tse, S. K., Lam, J. W. I., & Loh, E. K. Y. (2010). Does the gender of the teacher matter in the teaching of reading literacy? Teacher gender and pupil attainment in reading literacy in Hong Kong. *Teaching and Teacher Education*, 26, 754–759.
- Lerner, J. V., Lerner, R. M., & Zabski, S. (1985). Temperament and elementary school children's actual and rated academic performance: A test of "goodness-of-fit" model. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 26(1), 125–136.
- Lewis, M. (1989). Culture and biology: The role of temperament. In P. R. Zelazo & R. G. Barr (Eds.), *Challenges to developmental paradigms: Implications for theory, assessment, and treatment* (pp. 203–223). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist*, 45(4), 513–520.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Martin, R. P. (1989a). Activity level, distractibility, and persistence: Critical characteristics in early schooling. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 451–462). Chichester, UK: John Wiley & Sons, Ltd.
- Martin, R. P. (1989b). Temperament and education: Implications for underachievement and learning disabilities. In W. B. Carey & S. C. McDevitt (Eds.), *Clinical and educational applications of temperament research* (pp. 117–120). Berwyn, PA: Swets North America.
- Martin, R. P. (1992). Child temperament effects on special education: Process and outcomes. *Exceptionality*, 3(2), 99–115.
- Martin, R. P. (1994). Child temperament and common problems in schooling: hypotheses about causal connections. *Journal of School Psychology*, 32(2), 119–134.
- Martin, R. P., & Bridger, R. C. (1999). *The temperament assessment battery for children -revised: A tool for the assessment of temperamental traits and types of young children*. Unpublished manual.

- Martin, R. P., Drew, K. D., Gaddis, L. R., & Moseley, M. (1988). Prediction of elementary school achievement from preschool temperament: Three studies. *School Psychology Review, 17*(1), 125–137.
- Martin, R. P., & Holbrook, J. (1985). Relationship of temperament characteristics to the academic achievement of first-grade children. *Journal of Psychoeducational Assessment, 3*, 131–140.
- Martin, R. P., Nagle, R., & Paget, K. (1983). Relationships between temperament and classroom behavior, teacher attitudes, and academic achievement. *Journal of Psychoeducational Assessment, 1*, 377–386.
- Martin, R. P., Olejnik, S., & Gaddis, L. (1994). Is temperament an important contributor to schooling outcomes in elementary school? Modeling effects of temperament and scholastic ability on academic achievement. In W. B. Carey & S. C. McDevitt (Eds.), *Prevention and early intervention: Individual differences as risk factors for the mental health of children* (pp. 59–68). New York, NY: Brunner/Mazel.
- Maziade, M., Cote, R., Boutin, P., Boudreault, M., & Thivierge, J. (1986). The effect of temperament on longitudinal academic achievement in primary school. *Journal of the American Academy of Child Psychiatry, 25*, 692–696.
- Mendez, J. L., McDermott, P., & Fantuzzo, J. (2002). Identifying and promoting social competence with African American preschool children: Developmental and contextual considerations. *Psychology in the Schools, 39*(1), 111–123.
- Michaelowa, K. (2001). Primary education quality in francophone sub-Saharan Africa: Determinants of learning achievement and efficiency considerations. *World Development, 29*(10), 1699–1716.
- Newman, J., Noel, A., Chen, R., & Matsopoulos, A. S. (1998). Temperament, selected moderating variables and early reading achievement. *Journal of School Psychology, 36*(2), 215–232.
- Nigg, J. T., Goldsmith, H. H., & Sachek, J. (2004). Temperament and attention-deficit hyperactivity disorder: The development of a multiple pathway model. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 42–53.
- OECD. (2004). Learning for tomorrow's world: First results from PISA 2003. Paris: OECD.
- OECD. (2007a). PISA 2006 Science competencies for tomorrow's world: Volume 1, analysis. Paris: OECD.
- OECD. (2007b). PISA 2006 Science competencies for tomorrow's world: Volume 2, data. Paris: OECD.
- OECD. (2010). PISA 2009 Results: What students know and can do: Student performance in reading, mathematics and science (Volume I), OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264091450-en>
- Oliver, P. H., Guerin, D. W., & Gottfried, A. W. (2007). Temperamental task orientation: Relation to high school and college educational accomplishments. *Learning and Individual Differences, 17*, 220–230.
- Orth, L. C., & Martin, R. P. (1994). Interactive effects of student temperament and instruction method on classroom behavior and achievement. *Journal of School Psychology, 32*(2), 149–166.

- Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94*(2), 396–404.
- Posner, M. I., & Rothbart, M. K. (2007). *Educating the human brain*. Washington, DC: American Psychological Association.
- Pullis, M. E. (1989). Goodness of fit in classroom relationships. In W. B. Carey & S. C. McDevitt (Eds.), *Clinical and educational applications of temperament research* (pp. 117–120). Berwyn, PA: Swets North America.
- Pullis, M. E., & Cadwell, J. (1982). The influence of children's temperament characteristics on teachers' decision strategies. *American Educational Research Journal, 19*(2), 165–181.
- Radke-Yarrow, M., Richters, J., & Wilson, W. E. (1988). Child development in a network of relationships. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Relationships within families: Mutual influences* (pp. 48–67). Oxford University Press, NY: Clarendon Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). (Advanced quantitative techniques in the social sciences series No. 1). Thousand Oaks, CA: Sage .
- Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569–578.
- Rothbart, M. K. (1989). Biological processes in temperament. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 77–110). Chichester, UK: John Wiley & Sons, Ltd.
- Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16*(4), 207–212.
- Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, 6th ed., pp. 99–166). Hoboken, NJ: John Wiley & Sons.
- Rothbart, M. K., & Jones, L. B. (1998). Temperament, self-regulation, and education. *School Psychology Review, 27*(4), 479–491.
- Rudasill, K. M., Gallagher, K. C., & White, J. M. (2010). Temperamental attention and activity, classroom emotional support, and academic achievement in third grade. *Journal of School Psychology, 48*, 113–134.
- Rudasill, K. M., & Rimm-Kaufman, S. E. (2009). Teacher-child relationship quality: The roles of child temperament and teacher-child interactions. *Early Childhood Research Quarterly, 24*, 107–120.
- Ruus, V.-R., Veisson, M., Leino, M., Ots, L., Pallas, L., Sarv, E.-S., et al. (2007). Student's well-being, coping, academic success, and school climate. *Social Behavior & Personality, 35*, 919–936.
- Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16*(2), 125–141.
- Salvano-Pardieu, V., Fontaine, R., Bouazzaoui, B., & Florer, F. (2009). Teachers' sanction in the classroom: effect of age, experience, gender and academic context. *Teaching and Teacher Education, 25*, 1–11.

- Sanson, A. V., Smart, D. F., Prior, M., Oberklaid, F., & Pedlow, R. (1994). The structure of temperament from age 3 to 7 years: Age, sex, and sociodemographic influences. *Merrill-Palmer Quarterly*, *40*, 233–252.
- Schoen, M. J., & Nagle, R. J. (1994). Prediction of school readiness from kindergarten temperament scores. *Journal of School Psychology*, *32*(2), 135–147.
- Shin, S. H. (2003). Building evidence to promote educational competence of youth in foster care. *Child Welfare*, *132*(5), 615–632.
- Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology*, *43*, 39–60.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*(4), 323–355.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Skelton, C. (2002). The ‘feminisation of schooling’ or ‘re-masculinising primary education’? *International Studies in Sociology of Education*, *12*(1), 77–96.
- Slabach, E. H., Morrow, J., & Wachs, T. D. (1991). Questionnaire measurement of infant and child temperament: Current status and future directions. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives* (pp. 205–234). New York: Plenum.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290–312). Washington DC: American Sociological Association.
- Sokal, L., Katz, H., Chaszewski, L., & Wojcik, C. (2007). Good-bye, Mr. Chips: Male teacher shortages and boys’ reading achievement. *Sex Roles*, *56*, 651–659.
- Stevenson-Hinde, J. (1988). Individuals in relationships. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Relationships within families: Mutual influences* (pp. 68–80). Oxford University Press, NY: Clarendon Press.
- Stevenson-Hinde, J., & Hinde, R. A. (1986). Changes in associations between characteristics and interactions. In R. Plomin & J. Dunn (Eds.), *The study of temperament: Changes, continuities, and challenges* (pp. 115–129). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stipek, D. (2002). Communicating expectations. In D. Stipek (Ed.), *Motivation to learn: Integrating theory and practice* (pp. 210–229). Boston, MA: Allyn & Bacon.
- Strelau, J. (1998). *Temperament: A psychobiological perspective*. New York, NY: Plenum Press.
- Stuhlman, M. W., & Pianta, R. C. (2002). Teachers’ narratives about their relationships with children: Associations with behavior in classrooms. *School Psychology Review*, *31*(2), 148–163.
- Thomas, A., & Chess, S. (1977). *Temperament and development*. New York, NY: Brunner/Mazel.

- Tilastokeskus (2011). Suomen virallinen tilasto (SVT): *Koulutukseen hakeutuminen* [verkkojulkaisu] ISSN=1799-4500. 2010. Helsinki, Finland: Tilastokeskus [viitattu: 20.4.2012] Retrieved from http://www.stat.fi/til/khak/2010/khak_2010_2011-12-13_tie_001_fi.html.
- UNESCO. (2005). *EFA global monitoring report 2005: The quality imperative*. Paris: UNESCO.
- Van Houtte, M. (2007). Exploring teacher trust in technical/vocational secondary schools: male teachers' preference for girls. *Teaching and Teacher Education*, 23, 826–839.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behaviour. *Journal of Personality and Social Psychology*, 95(5), 1202–1216.
- Väljjärvi, J., Kupari, P., Linnakylä, P., Reinikainen, P., Sulkunen, S., Törnroos, J., & Arffman, I. (2007). *The Finnish success in PISA—and some reasons behind it 2. PISA 2003*. Jyväskylä, Finland: Kirjapaino Oma Oy.
- Walker, S., Berthelsen, D., & Irving, K. (2001). Temperament and peer acceptance in early childhood: Sex and social status differences. *Child Study Journal*, 31(3), 177–192.
- Warwick, D. P., & Jatoi, H. (1994). Teacher gender and student achievement in Pakistan. *Comparative Education Review*, 38(3), 377–399.
- Windle, M. (1992). Revised dimensions of temperament survey (DOTS-R): Simultaneous group confirmatory factor analysis for adolescent gender groups. *Psychological Assessment*, 4(2), 228–234.
- Windle, M., Hooker, K., Lerner, K., East, P. L., Lerner, J. V., & Lerner, R. M. (1986). Temperament, perceived competence, and depression in early and late adolescents. *Developmental Psychology*, 22(3), 384–392.
- Windle, M., Iwawaki, S., & Lerner, R. M. (1988). Cross-cultural comparability of temperament among Japanese and American preschool children. *International Journal of Psychology*, 23, 547–567.
- Windle, M., & Lerner, R. M. (1986). Reassessing the dimensions of temperamental individuality across the life span: The revised dimensions of temperament survey (DOTS-R). *Journal of Adolescent Research*, 1(2), 213–230.
- Yang, K. (1986). Chinese personality and its change. In M. H. Bond (Ed.), *The Psychology of the Chinese People* (pp. 106–170). New York, NY: Oxford University Press.
- Zhou, Q., Lengua, L. J., & Wang, Y. (2009). The relations of temperament reactivity and effortful control to children's adjustment problems in China and the United States. *Developmental Psychology*, 45(3), 724–739.
- Zhou, Q., Main, A., & Wang, Y. (2010). The relations of temperamental effortful control and anger/frustration to Chinese children's academic achievement and social adjustment: A longitudinal study. *Journal of Educational Psychology*, 102(1), 180–196.