University of Trier

Faculty I – Department of Psychology

Giftedness Research and Education

Diagnostic Competencies of Teachers

Accuracy of Judgment, Sources of Bias, and Consequences of (Mis-)Judgment

Dissertation submitted to the University of Trier, Faculty I – Department of Psychology to obtain the degree of doctor rerum naturalium (Dr. rer. nat.) in Psychology.

Supervisors:

Prof. Dr. Franzis Preckel

University of Trier

Prof. Dr. Ricarda Steinmayr

TU Dortmund University

submitted by:

Dipl.-Psych. Rachel Wollschläger

on September 20, 2016

To those who encouraged and guided me throughout my entire journey,
I am much obliged.

Abstract

Educational assessment tends to rely on more or less standardized tests, teacher judgments, and observations. Although teachers spend approximately half of their professional conduct in assessment-related activities, most of them enter their professional life unprepared, as classroom assessment is often not part of their educational training. Since teacher judgments matter for the educational development of students, the judgments should be up to a high standard. The present dissertation comprises three studies focusing on accuracy of teacher judgments (Study 1), consequences of (mis-)judgment regarding teacher nomination for gifted programming (Study 2) and teacher recommendations for secondary school tracks (Study 3), and individual student characteristics that impact and potentially bias teacher judgment (Studies 1 through 3). All studies were designed to contribute to a further understanding of classroom assessment skills of teachers. Overall, the results implied that, teacher judgment of cognitive ability was an important constant for teacher nominations and recommendations but lacked accuracy. Furthermore, teacher judgments of various traits and school achievement were substantially related to social background variables, especially the parents' educational background. However, multivariate analysis showed social background variables to impact nomination and recommendation only marginally if at all. All results indicated differentiated but potentially biased teacher judgments to impact their far-reaching referral decisions directly, while the influence of social background on the referral decisions itself seems mediated. Implications regarding further research practices and educational assessment strategies are discussed. The implications on the needs of teachers to be educated on judgment and educational assessment are of particular interest and importance.

Table of Contents

List of Abbreviations

ADHD	attention deficit/hyperactivity disorder
ANOVA	analysis of variance
CFA	confirmatory factor analysis
CFI	comparative fit index
EFA	explorative factor analysis
FEESS	Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen
Gf	fluid intelligence
GPA	grade-point average
IEA	International Association for the Evaluation of Educational Achievement
ICC	intra-class correlation
IQ	Intelligence Quotient
MLR	Restricted Maximum Likelihood
MI	measurement invariance
OECD	Organization for Economic Co-operation and Development
PIRLS	Progress in International Reading Literacy Studies
PISA	Program for International Student Assessment
S1	Sample 1
S2	Sample 2
SEE	standard error of estimation
SEM	Schoolwide Enrichment Model
SES	socio-economic status
T(H)INK	Test for (Highly) Intelligent Kids
THINK 1–4	Test for the Assessment of Intelligence in Childhood
TJ	teacher judgment
TJA	teacher judgment accuracy
Tukey's HSD	Tukey's honest significance difference
USA	United States of America

List of Figures

List of Tables

1. Introduction

"A knowledgeable teacher
is the foundation of informed assessment."
(Wolf, 1993, p. 518)

According to Shulman (1998), "all professions are characterized by the following attributes: the obligations of *service* to others, as in a 'calling', *understanding* of a scholarly or theoretical kind, a domain of skilled performance or *practice*, the exercise of *judgment* under conditions of unavoidable uncertainty, the need for *learning from experience* as theory and practice interact, and a professional *community* to monitor quality and aggregate knowledge" (p. 516). Transferring this understanding of a profession to the teaching staff, teachers should educate others for the purpose of enhancing learners' educational development, have profound knowledge of the taught subject, correctly apply suitable methods of teaching, make ad hoc decisions in complex environments such as class rooms, reflect their professional conduct and be part of professional teams such as the teaching staff of a school or regional departments of their subject of study.

Kunter and Klusemann (2010) searched for the "competent teacher" and integrated several characteristics that have been found to be important into an interactive model. They compared teachers regarding subject knowledge, professional motivation and strategies of self-regulation, and identified three latent classes of teachers. Although, the role model and the self-regulator types differed regarding self-regulatory competencies and efforts in continuous training, both types were found to be motivated, well-educated and competent in teacher-student interaction. The third type, however, caused concern, as it was characterized by lacking subject knowledge, applying destructive teaching methods, having internalized destructive motivational mindsets, and showing the lowest engagement in further educational training. Although results are dependent on the sample, it is noteworthy, that the majority belonged to the last, problematic category. Regardless of the teachers' level of competence, his or her judgment is of consequence and has an undeniable impact on students' educational development and with that his or her professional future (Fischbach, Baudson, Preckel, Martin & Brunner, 2013).

The present dissertation focuses on diagnostic competencies of teachers and comprises three research studies accompanied by a common theory and a general discussion. The theoretical background, Chapter 2, addresses educational and psychological assessment as well as teacher diagnostic abilities and responsibilities. The research question is explicated in Chapter 3. Chapters 4 to 6 present the three research articles, each structured in theoretical background, present study with research questions, methods, results, and discussion. In Chapter 7, the results of the three studies are discussed regarding their contribution towards the further understanding of accuracy of teacher judgment, sources of bias in judgment, and the consequences of (mis-)judgment. After addressing strengths and limitations of the present dissertation, implications for further research and practical consequences are presented. Finally, Chapter 8 summarizes the essence of the present dissertation.

Educational assessment has occupied the center stage of pedagogic-psychological research in Germany, especially since the Program for International Student Assessment (PISA) study of 2000 has pointed into the direction of social disparities in judgment and lacking diagnostic abilities in teachers (OECD, 2001). The aim of the present dissertation is to contribute to a further understanding of classroom assessment skills of teachers.

2. Theoretical background

This chapter addresses the theoretical background on which all three research articles are based. Teacher judgments and the assessment of school achievement are a way of operationalizing educational assessment. Assessment of any kind, but educational assessment in particular, needs preceding knowledge regarding its correct application. Therefore, after introducing educational assessment, theories of and findings on professional knowledge of teachers and their diagnostic competencies are presented. In conclusion, the relevance of educational assessment in general and in the light of the present dissertation are highlighted.

2.1. Educational assessment

According to Ingenkamp and Lissmann (2008) educational assessment is mostly based on observation that aims at comparing, analyzing, predicting and interpreting human behavior within the educational setting. If a teacher would, for example, observe a student outperforming his former achievement, the educator would analyze possible reasons for this change to predict the student's future performance and interpret available information to come to a judgment. While psychological assessment typically consists of psychometrically evaluated measures that aim at measuring a person's trait or ability, educational assessment rather relies on tests of different degrees of standardization, teacher judgments, systematic and unsystematic observations to assess students' academic achievement (Tent & Stelzl, 1993).

Leutner (2010) differentiates educational assessment into formal and informal assessment types. The formal test is characterized by a construction on test theoretical background and a high level of standardization regarding application, evaluation and interpretation. It aims at meeting the requirements of objectivity, reliability and validity of psychometrically acceptable tests. The informal tests are rather designed on demands of the teachers' learning progress in class and his or her class specific needs of assessment, rather than on test theoretical standards. Although they lack standardization, informal tests are not

automatically violating quality criteria of testing (ibid.). Typical assessment tools applied by teachers are "oral questioning of students, observation, written work products, oral presentations, interviews, projects, portfolios, tests, and quizzes" (Shepard, Hammerness, Darling-Hammond & Rust, 2005, p. 294).

Results from the Progress in International Reading Literacy Studies (PIRLS) by the International Association for the Evaluation of Educational Achievement (IEA) revealed, 51% of German teachers rely on their own professional judgment, when assessing their students' reading literacy. Their professional judgment was accompanied by classroom tests (38%) and diagnostic tests (23%). Only 11% named national or regional achievement tests to be important for their assessment of reading literacy (Mullis, Martin, Kennedy & Foy, 2007, p. 238).

2.1.1. Foci of educational assessment

Educational assessment has various purposes that are hardly assimilable: Based on their performance, students are selected for educational paths, socialized and disciplined for the societal requirements, their achievement development is monitored, and their future achievement is prognosticated. Furthermore, teaching methods or even schooling systems are possibly adjusted based on the results of educational assessment(Sacher, 2009). Education itself aims at socializing and qualifying students for requirements later in life and selecting those who have prospectively high potential for success and because of their potential success are of societal value. Through educational assessment those aims are operationalized and legitimized (Fischbach, 2013). Biesta (2015) distinguished three domains of educational purposes: Qualification, as in acquiring knowledge and skills, socialization, as in learning about social norms, values and tradition, and subjectification, as in personality development. Although it is genuinely assumed that the main purpose of institutionalized education is to teach the learner knowledge and skills for their own personal(-ality) development, the societal function of education may be considered far more important (Fend, 1980, 2008; Luhmann, 2002).

Society targets at bringing up children that secure the societal survival by integrating themselves into the society through education (Bourdieu & Passeron, 1971, 2007). By teaching values, beliefs, knowledge, and skills that are held important for the society, societal structures aspire to stand the test of time. As Fischbach (2013) has concluded, the dilemma of society- versus student-centered assessment is difficult to overcome. Teachers need to identify strengths and weaknesses of their individual students through assessment,

adjust their teaching in class based on the results of their assessment and select the "right" students for educational paths aiming at societal parts by summing up their performance in final grades. The results of any kind of assessment, however, are dependent of the comparison standard it is held against.

2.1.2. Frames of reference in educational assessment

To decide whether a student performed well or poorly, a norm is needed to define "good" and "bad" performance. Achievement of all kinds may be compared against different standards. If a students' performance is related to the performance of a social group, e. g., the students' class, his or her performance may be considered good, if performing at least on the average level of the group. This frame of reference is called the *social norm*. Comparing a students' performance to his or her former achievement, the *individual norm* is applied and stability of or improvement in the performance level is considered good. The third norm is the *criterion norm*, for which previously defined criteria should be met in order for performance to be good (Sacher, 2009). Although all three norms have their entitlement in educational assessment, the criterion norm should be mainly focused upon when judging performance (ibid.). In fact, federal regulations are violated, if considering the individual or social norm when grading (KMK, 1968). However, teacher judgments of various students' traits are often biased by context factors (i. e., the class). A case in point is class-average ability or achievement, which acts as a frame of reference against which individual students are evaluated (Marsh, 1984, 1987; Marsh, Trautwein, Lüdtke & Köller, 2008). Effects of the class-average ability or achievement on teacher judgments are called reference group effects. Trautwein and Baeriswyl (2007) examined the influence of class-average achievement on teacher recommendations for academic tracks in Switzerland. Using a standardized test to measure students' achievement, they were able to show a negative influence of the class-average achievement not only for teacher recommendations regarding future academic tracks but also on teacher estimates of students' cognitive ability and grade-point average (GPA). Wollschläger, Baudson, Schmitt, Fendahl, and Preckel (2014) replicated the negative impact of the class-average cognitive ability on teacher judgment (TJ) of their students' cognitive ability and academic motivation, as well as the students' GPA in two independent samples. Both studies hint toward a violation of grading standards that affect students' educational careers directly (Trautwein & Baeriswyl, 2007) and indirectly through the students' grades (Wollschläger et al., 2014).

2.2. German schooling system

The data collection of the present dissertation took place in Germany. In Germany, the federal states are responsible for their own education program, their schools and schooling standards. Therefore, only those foundation pillars applicable to almost all states will be sketched out in this section. In general, all children in Germany are required to attend primary school when they turn six years old. Some exclusions are possible, for example starting school accelerated at five years or delayed at seven years. For all children living in Germany school attendance is obligatory. Primary school usually lasts four years. By the end of primary school, teachers recommend a type of secondary school, but in most federal states of Germany the final decision where to enroll the student rests on the students' parents. The highest track – the "Gymnasium" – lasts eight to nine years and allows students to pursue the A-level exams directly and with it attend university. Lower track secondary schools may are further divided into two tracks: the "Realschule", which lasts six years, and the "Hauptschule", which lasts five years. Both aim at qualifying students for the apprenticeships/vocational training. Another type of secondary school is the integrated school, the "Gesamtschule", which usually offers all tracks parallel allowing students to switch between tracks.

Recent political decisions and educational developments have led to changes in the variety of secondary schools. The systems developed from a strict three-tier secondary school system towards a two-tier system, with the "Gymnasium" as the highest track and the "Realschule" as the most common secondary school track (Bellenberg, 2012). In general, students tend to remain in the secondary track in which they have started (Bellenberg, 2012; Lohmann & Groh-Samberg, 2010) and, although students are able to switch between tracks in both directions, it is four times more likely to change to a lower track, than to a higher track (Berkemeyer, Bos, Manituis, Hermstein & Khalatbari, 2013).

2.3. Professional knowledge of teachers

No matter where and what teachers teach, there are some fundamental similarities regarding the professional knowledge required to be capable of teaching. According to Shulman (1986, 1987), teachers' professional knowledge may be distinguished into content knowledge referring to the subject taught, pedagogical content knowledge referring to pedagogical specifics of the subject taught, and general pedagogical knowledge referring to basic knowledge of theoretical and applied pedagogy. While content and pedagogical content

knowledge has been of continuous importance in educational research, general pedagogical knowledge has been rather unobserved (Voss, 2010). Voss, Kunter, and Baumert (2011) defined general pedagogical/psychological knowledge as "knowledge needed to create and optimize teaching-learning situations" (p. 952). Within their theoretical framework, general pedagogical/psychological knowledge is comprised of the facets classroom management, teaching methods, and classroom assessment and students' heterogeneity (ibid.). The knowledge of classroom assessment includes the development, application, evaluation and interpretation of assessment-related activities. Stiggins (1991) introduced the term of "assessment literacy", meaning correct application of assessment tools combining qualitatively high standards and careful evaluation of the produced data. Additionally, according to Stiggins (2007) educational assessment should lead to a reflection of learning goals. Classroom assessment[1] is "crucial in enabling teachers to judge students' progress toward their goals and in helping them to adapt their instruction to the individual needs of their students" (Voss et al., 2011, p. 953). This also implies, that results of any assessment should not stand alone but should generally be accompanied by intensive reflection of the teachers' own teaching methods and learning goals. The importance of classroom assessment in teachers' profession becomes obvious, when bearing in mind that teachers spend nearly half of their professional time in activities related to educational assessment (Plake & Impara, 1997).

2.3.1. Classroom assessment in teacher education

The assessment of the student's ability or learning process is part of every teachers' professional career. Schafer (1993), as well as Plake and Impara (1997), found that the majority of teachers were not prepared for assessment through their teacher education. Biethahn and Hogrebe (2016) asked 104 primary school teachers to name their sources of knowledge regarding grading and other assessment-related activities. Only 16% stated that educational assessment was addressed in courses when studying at the University and of all their acquired knowledge only, 4% comes from teacher education on a University level. The majority of knowledge was acquired through colleagues, self-instruction, and within on the job training. All in all, teachers seem to lack knowledge on educational assessment as a particular part of teachers' general pedagogical/psychological knowledge, at the same time teacher education seem to lack related content as well.

[1]Because the present dissertation focuses on educational assessment through teachers, all other facets are not further described or discussed, for further details and a critical appraisal of general pedagogical/psychological knowledge see Voss, Kunina-Habenicht, Hoehne, and Kunter (2015).

2.4. Diagnostic competencies of teachers

While classroom assessment refers to a specific skill that is considered to be essential to a teachers' profession, diagnostic abilities of teachers refers to broader skills. Teachers' diagnostic competencies are usually understood as the ability to judge students' potential, knowledge, achievement, and other school-related constructs, integrating that information and drawing informed conclusions on teaching methods of all kinds (Schrader, 2006). Furthermore, teachers should be capable of assessing and evaluating students' learning to foster each student's potential by balancing student ability and task difficulty (Baumert & Kunter, 2006). Although the students' individual intellectual ability has been found to be one of the most important single predictors of the students' academic achievement, it is not solely dependent on it (Neisser et al., 1996, Roth et al., 2015). Academic achievement is rather a conglomerate of individual characteristics such as: cognitive ability, learning activity, learning/achievement motivation, self-regulation, learning strategies, and context characteristics such as family background, student-teacher interaction, school and teaching environment (e. g., Weinert, 2001; Helmke, Rindermann & Schrader, 2008). If a teacher is asked to estimate the students' academic potential, all those characteristics should be considered.

2.4.1. Teacher judgments

Grading students' achievement is a kind of TJ and presents formative and summative assessment at the same time: Teachers grade their students' work continuously over the school year and give out a final grade at the end of the year, representing the students' overall achievement. TJs of both their students' cognitive ability and their academic achievement have been found to be rather accurate overall, but also vary greatly. For instance, Hoge and Coladarci (1989) identified correlations from $r = .28 - .92$ ($Mdr = .66$) between TJs of student performance and students' actual performance in their analysis of 16 studies. In their meta-analysis over 75 studies, Südkamp, Kaiser, and Möller (2012) reported a median correlation of $r = .63$ between teacher-rated and actual achievement. Grading activities tend to refer to a restricted frame of reference, which usually is social instead of the criterion comparison standard, such as the taught class (e. g., Baudson, Fischbach & Preckel, 2014).

Regarding the judgment of students' traits and behavior, Anders, McElvany, and Baumert (2010) investigated the degree of differentiation with which teachers judge their students.

In their study teachers judged their students on a variety of constructs and traits related to academic achievement. Through explorative factor analysis (EFA) they extracted three dimensions; (1) *talent and achievement* – composed of cognitive ability, psychological robustness, and academic skills, (2) *social abilities and behavior* – composed of the student's social behavior in class and impulse control, (3) and *motivation and learning virtue* – composed of learning/achievement motivation, effort, and discipline. It is important to point out, that the judgment of a student's traits is at least related, but possibly intertwined with the judgment of students' academic achievement vice versa. Regarding the facets extracted by Anders et al. (2010), the judgments of talent and achievement were highly related to grades in Math ($r = .75$) and German ($r = .77$), those of motivation and learning virtue were highly related as well (Math: $r = .60$; German: $r = .67$), while the relation of social abilities and behavior and grades were of medium height (Math: $r = .34$; German: $r = .43$).

2.4.2. Inter- and intra-rater agreement of teacher judgments

The quality of TJs has been discussed controversially for decades and across various disciplines (see Südkamp et al., 2012, for details). In some studies, inter-rater agreement and inter-rater reliability is used synonymously, but their meaning differs: If teachers judged three papers, the reliability would be high if the papers would be judged with A, B, C by one teacher and D, E, F by the other, but there would be no agreement between assigned grades. While agreement refers to the absolute match in value/grade, reliability usually refers to the relative order of TJs. The reported studies comparing judgment *between* teacher, investigated *inter*-rater agreement, while those comparing judgment of *one teacher over time*, focused on the *intra*-rater reliability.

Starch and Elliot (1912) compared grading standards for student papers in English between leading English teachers of various schools by sending out two ungraded and uncommented student papers and asking the teachers to return them corrected and graded. Out of 200 student papers, 142 were returned graded and included in further analyses. Grades varied from 64 (failing) to 98 (passing with excellence) on the 100-point grading scale for paper one and from 50 (failing) to 98 (passing with excellence) for paper two, hinting at great inter-individual differences. When conducting a similar study for math, Starch and Elliot (1913) found grades to vary from 28 (failing) to 92 (passing with excellence) points, underlining previous findings for English. Brimi (2011) replicated those findings nearly one hundred years later, when asking 73 teachers to grade school work in

English and found grades ranging from 50 (failing) to 96 (passing with excellence), even after intensive training on grading essays. Hartog and Rhodes (1936) compared teacher judgments regarding oral exams and found similar results. Four decades later, Birkel and Pritz (1980) recorded an oral exam in geography that was graded "satisfactory" and re-enacted it with varied speech tempo. Teachers graded either the slower (21 minutes) or the faster (16 minutes) version. Grades ranged from 1 (excellent) to 5 (failed), hinting at differences between teachers again. Moreover, the faster version was judged one grade point better than the slower version on average, although they were exactly the same regarding content, language, gestures and facial expressions. To summarize, great variation in teacher judgment does not seem to be restricted to specific subjects nor forms of assessment.

Eells (1930) compared teachers' intra-individual judgment. After a lapse of eleven weeks, 61 grammar school teachers were asked to regrade the written exams in geography and history. He found reliability coefficients ranging from .25 to .51, leading him to conclude that "the variability of grading is about as great in the same individual as in groups of different individuals" (p.48). Macnamara and Madaus (1969) found that grades did not differ more between teachers than within one teacher in exams at the End of Irish grammar schools, and replicated Eells (1930) findings.

All in all, TJ lacks reliability and agreement between and within teachers. Judgments seem exposed to subjective grading standards and techniques, which could be based on the lack of educating teachers on educational assessment (see also Daniel & King, 1998). Nevertheless, it should be kept in mind, that the reported findings are tendentially old. However, findings regarding the (lacking) validity of teacher judgment are most present and will be outlined in the following section.

2.4.3. Interferences on teacher judgment

Even though teachers had no additional information on the students, Starch and Elliot (1912; 1913), Eells (1930), and Brimi (2011) found great variations in TJs. In the school-house setting, teachers do not grade the mere anonymous performance on a task, but have background information on student characteristics and his or her prior school achievement. Influences on and biases in TJs have been a prominent research field within educational and psychological research. If actual differences in academic or social abilities or motivation exist, differences in judgment would be correct. However, if social background variables lead to lower or less favorable teacher judgments even if the objective

level is comparable, any differences in judgment must be considered as biased and invalid as a consequence.

Social inequality results in social injustice, if differences regarding socially relevant factors such as income, education, and health, violate societal conceptions of justice (Diewald & Riemann, 2014). In Article 26 of the Universal Declaration of Human Rights, the right to education is expressed. Education should be equally accessible to everyone regardless his or her gender, race, religion, ethnicity and/or socio-economic status (The United Nations, 1948). Nevertheless, results of the PISA study from 2000 indicated that the relation between educational success and socio-economic background was closest in Germany (OECD, 2001). Not only the high interdependence of students' background variables and their educational paths, but also the overall performance of German students in an average level in comparison to other countries, led to establishment of research projects aiming at understanding the mechanism of the German school system and the social disparities. One focus was set on the interference of TJs and students' background variables (e. g., Hörstermann, Krolak-Schwerdt & Fischbach, 2010).

Most prominent sources of interference with TJs are socio-economic status or any proxy of the latter, students' gender[2] and migration background (Gniewosz & Gräsel, 2011). As all studies considered various social background variables and their impact on TJs of school-related traits, school achievement and educational path related recommendations, the particularly findings on the above-mentioned sources of interference will be presented in each article (see Chapters 4 to 6).

2.5. Relevance of educational assessment

TJs have a high impact on the students' educational career. Especially transition habits in an educational system hint at the importance of the TJ of (potential) academic ability. After four years of primary school, German students are tracked into secondary school – the tracking recommendation is based on the students' grades, which are the results of teachers' classroom assessment practices and a TJ par excellence. When graduating from secondary school, students in Germany are faced with universities and national entrance boards considering only the final GPA when selecting their future university students.

[2]When referring to gender, the biological sex is meant. Although social gender and biological sex are known to be distinct, an anonymous reviewer asked for the replacement of sex with gender. Therefore, the students' or teachers' sex was replaced with the students' or teachers' gender in all studies for reasons of consistency.

Even for vocational trainings candidates are chosen on basis of their secondary or vocational school grades. Fischbach et al. (2013) predicted life outcomes at the age of 52 through the TJs of students' cognitive ability at the age of 12. They found TJs to predict educational attainment, socioeconomic achievement and health pointing towards the direction of self-fulfilling prophecies: If a teacher thinks a student to be intellectually capable, he or she may perform even better (in later life) than could be expected from standardized test results (at the time of the judgment; ibid.). What teachers believe a child is capable of, influences his or her choice of instructions manners (McElvany et al., 2009, Baumert & Kunter, 2006). When a child's cognitive ability is underestimated, lower expectations may be risen, leading to the selection of inadequate, i. g. too simple, instruction materials by the teacher, and, eventually, to potential remaining underdeveloped. When controlling for actual performance, relations between judgment and outcome decreased, hence, Fischbach et al. (2013) discuss TJs of a student trait to reflect a conglomerate of students' attributes and school achievement, rather that the judged trait alone.

To conclude, educational assessment in general and TJs in particular are of practical relevance for student development, therefore teachers should judge and grade as precise, unbiased and valid regarding the judged content as possible. To be a competent teacher, evolved classroom assessment skills are necessary. Previous research has shown that TJ in general is rather lacking precision, either because of the choice and construction of unsuitable methods or because of biasing influences. However, teachers are not solely to blame as the majority entered the professional life unprepared because of missing thematization of and practice in educational assessment in their education.

3. The present studies

As outlined before, diagnostic competencies of teachers, are defined by their ability to judge students adequately concerning ability, achievement and other school-related constructs (Schrader, 2006). Because teachers observe students in various and differently challenging settings over a long period of time, they gather unique, possibly contradicting information. Based on the information they gather, teachers form implicit and explicit opinions and expectations, which affect their judgments (Wyatt-Smit, Castleton, Freebody & Cooksey, 2003). Their expectations and judgments of a student's abilities have significant influence on students' achievement and thus career and success in life (Südkamp et al., 2012; Fischbach et al., 2013). Thus, their diagnostic abilities in general and classroom assessment skills in particular are of high practical relevance – for student development in general, for nominating students for gifted education, as well as for recommending the suitable secondary track.

Various national and international student assessment studies have shown the continuous importance of the students' social origin for educational paths and success in the German educational system (Maaz, Baumert, Gresch & McElvany, 2010; Jonkmann, Maaz, Neumann & Gresch, 2010). Therefore, student and family background variables were included in all of the three enclosed studies to investigate the continuous impact of a students' origin – 16 years after the first PISA-shockwaves kicked off countless studies on and educational interventions for an educational system more just for all students.

The first study (entitled "Accuracy of teacher judgments of primary school students' cognitive ability", see Chapter 4) investigates the accuracy of TJ of students' cognitive ability following a long tradition of educational research. When it comes to judging their students' cognitive ability, teachers tend to equate cognitive potential to academic achievement (Hany, 1997). Instead of judging the students' potential, teachers tend to judge the students' performance or achievement (Baudson et al., 2014), but achievement does not equate potential perfectly (e. g., Hanses & Rost, 1998). Because teacher judgment accuracy (TJA) has been studied for a long time, especially within the field of gifted identification (e. g., Gear, 1979; Feldhusen, Asher & Hoover, 1984; Renzulli & Delcourt,

1986), various strategies to operationalize accuracy exist, but they have not been tested against each other yet. The present study aims at shedding further light on primary school TJA when judging the students' cognitive ability, variables biasing their judgment, and differences in results due to different methodological approaches to TJA.

The second study (entitled "Keeping the Gate: Components of Primary School Teachers' Nominations for a Gifted Program in Germany", see Chapter 5) focuses on components influencing teacher nominations for the participation in a pull out enrichment program for gifted students. Furthermore, differences in nominations between boys and girls concerning the components influencing the probability of the nomination are investigated. Within the conception of any program in gifted education, one crucial element is the successful acquisition of students whose needs and abilities match the offerings and requirements of the program (Mönks & Heller, 1994). Frequently, teachers are asked to nominate students for gifted programs (Baudson, 2010). Overall, the present study emphasizes the continuous importance of contextual influences and differences between girls and boys on the probability of being nominated for a gifted program.

The third and final study (entitled "Predictive value of students' social origin and teacher judgments for secondary school tracking decisions", see Chapter 6) examines teachers' secondary school track recommendation, as well as the parental final enrollment decision, within the highly tracked secondary schooling system in Germany. The teacher's recommendation, as well as the parental final enrollment decision, are a matter of importance for students' educational career. The aim of this paper is to analyze the impact of a student's social origin on tracking recommendations, school achievement, and acTJs, and how the last two influence tracking recommendation themselves. Furthermore, the relation between teacher track recommendation and parental enrollment decision is examined.

All three studies focus on the diagnostic competencies of German primary school teachers. Data was collected in the context of the piloting and standardizations studies of the Test for Assessment of Intelligence in Childhood (piloting version: Test for (Highly) Intelligent Kinds, T(H)INK, Baudson & Preckel, 2013a; standardized version: Test for Assessment of Intelligence in Childhood, THINK 1–4, Baudson, Wollschläger & Preckel, 2016). These field studies examined teachers, students, and the students' parents in their natural habitat providing naturalistic data. Table 3.1 provides an overview of the operationalization of the present studies and which field within the research on diagnostic competencies of teachers they tackle – judgment accuracy, sources of bias in judgment and consequences of (mis-)judgment.

Table 3.1.: General overview of the present studies, adapted from Fischbach (2013)

	Study 1	***Study 2***	***Study 3***
	Teacher judgment accuracy	***Teacher nomination for gifted programming***	***Secondary track recommendation and final enrollment***
Criterion	Students' IQ	Probability of nomination	Secondary track recommendation
Teacher related predictors	TJs of their students' cognitive ability, academic motivation and creativity	TJs of their students' cognitive ability, academic motivation and creativity	TJs of their students' cognitive ability, academic motivation and social behavior
Student related predictors	students' individual and family background variables	students' individual background variables	students' individual and family background variables, as well as indicators of previous educational history
Sample(s)	***Sample 1*** 679 primary school students Grade 2: $n = 354$ Grade 3: $n = 325$ ***Sample 2*** 2079 primary school students Grade 1: $n = 336$ Grade 2: $n = 428$ Grade 3: $n = 394$ Grade 4: $n = 921$	679 primary school students Grade 2: $n = 354$ Grade 3: $n = 325$	597 primary school students Grade 4: $N = 597$
Diagnostic competencies of teachers			
Accuracy	X		
Sources of bias	X	X	X
Consequences of (mis-)judgment		X	X

4. Accuracy of teacher judgments of primary school students' cognitive ability

Abstract. When judging students' cognitive ability, teachers tend to rate performance or achievement rather than potential. As potential and achievement are not perfectly related, this bears the risk of demanding too much or too little from students. The purpose of this study was to shed further light on primary school teacher judgment accuracy when judging the students' cognitive ability and variables biasing their judgment. Furthermore, we compared different methodological approaches to TJA. We analyzed data from two independent German samples. Sample 1 was composed of 679 primary school children and their teachers ($N = 46$). For Sample 2, we assessed 2,079 primary school children along with their 153 teachers. We applied correlational and regression-based approaches to TJA. Results showed that TJA rates depended on the approach applied. While TJs of cognitive ability correlated significantly and highly with tested Intelligence Quotient (IQ), further analyses showed them to be off by 12 IQ-points on average. Additionally, TJA was found to be low when judging heterogeneity in class but high when judging rank order and overall level of cognitive ability in their class. Family educational background and the students' mother tongue were related to TJs, i. e., the higher the family educational background the higher the teacher judged cognitive ability and German native speakers were judged higher in their cognitive ability. Accuracy rates differed by method; therefore, the method should be chosen attentively. Considering the substantial influence teachers have on student achievement and thus success in later life, TJA is of high practical relevance to student development and should be improved through teacher training.

4.1. Theoretical background

Teachers' diagnostic competencies are usually understood as the ability to judge students' potential, knowledge, achievement, and other school-related constructs, integrating that information and drawing informed conclusions on teaching methods and professional conduct (Schrader, 2006). The judgment of achievement-related behavior and traits presents an essential part of teachers' diagnostic competencies (Artelt & Gräsel, 2009). When judging their students' cognitive ability, teachers rate students' academic performance rather than their cognitive ability, i. e., potential (Baudson et al., 2014; Hany, 1997), although performance is no perfect indicator of potential (Hanses & Rost, 1998). Because accurate assessment is a prerequisite to adequate instruction, misjudgments of students' abilities may hamper the actualization of potential (McElvany et al., 2009). This is exemplified in teacher expectations, i. e., what teachers believe a child is capable of (Schrader & Helmke, 2001; Urhahne & Zhu, 2015; Valdez, 2013). Consequently, underestimating a child's cognitive ability may promote lower expectations. These lower expectations might lead to lower confidence and motivation in the child, selection of inadequate instruction materials by the teacher, and, eventually, to potential remaining underdeveloped. In contrast, mild overestimations of a child's cognitive ability may lead students to perform higher than what could be expected by potential alone (Baudson, 2011). Because teachers influence students' achievement and thus career/success in later life strongly (Fischbach et al., 2013), TJs are of high practical relevance to student development.

Following a long tradition of educational research, the present study examined TJA of students' cognitive ability and variables biasing their judgment. Through building on published work, this study aims to replicate and extend findings by combining different operationalization of accuracy. Those different operationalization are compared regarding their results in TJA, which to our best knowledge has not been done before.

4.1.1. Cognitive ability in the schoolhouse

Intelligence is usually understood as a general mental capability that is expressed by different kinds of abilities, such as the ability to learn (Gottfredson, 1997, Nisbett et al., 2012). A student's cognitive ability is important for his or her educational success (e. g., Spniath, Spinath, Harlaar & Plomin, 2006). Valerius and Sparfeldt (2015) investigated the relation of intelligence and school achievement. Through a nested-factor-model, the authors related a general and three specific (verbal, numerical, figural) intelligence factors

to a general and two specific (verbal, mathematical/scientific) achievement factors and found a substantial correlation of the general factors ($r = .44$). While verbal and numerical intelligence were more closely related to corresponding achievement factor, figural intelligence was related to both specific achievement factors.

Students' cognitive ability has been found to be related to grading practices, too. Kaiser, Möller, Helm, and Kunter (2015) presented achievement-related and achievement-unrelated information to trainee teachers and asked them to grade students' math performance. The authors found that information on the students' intelligence impacted the resulting math grade, i. e., the more intelligent the student, the higher the grade. While Kaiser et al. (2015) explicitly presented information on the students' cognitive ability, teachers usually do not know their students' intelligence, but infer cognitive ability through the proxy of students' performance and behavior (see also Renzulli & Delcourt, 1986). Because TJs of achievement and their accurate assumptions of students' potential are relevant to the students' educational success, judgments of cognitive ability should be accurate.

4.1.2. Accuracy of teacher judgments

Usually, TJA is examined by correlating teacher judgments (e. g., of students' cognitive ability) to objective test results (e. g., IQ) or to students' self-reported information (e. g., self-rated intelligence). Research suggests that TJs of both their students' cognitive ability and academic achievement are considered to be rather accurate overall, but also vary greatly between individuals (e. g., Schrader & Helmke, 2001). In their meta-analysis over 75 studies, Südkamp et al. (2012) reported a mean effect size of $r = .63$ between teacher-judged and actual achievement. DeYoung (2008) reported correlations of teacher-judged student intelligence and tested IQ between $r = .45$ and .80. Machts, Kaiser, Schmidt, and Möller (2016) reported a mean correlation of TJs and measures of diverse cognitive abilities of $r = .45$ in their meta-analysis of 106 effect sizes from 33 studies. They included judgments of intelligence, i. e., general ability ($r = .50$), cognitive ability ($r = .42$), giftedness ($r = .36$), and creativity ($r = .34$), revealing moderate effects. We refer to the correlation between teacher judgment and objective measure as *general accuracy* because it is the most common accuracy measure (e. g., DeYoung, 2008; Machts et al., 2016; Südkamp et al., 2012).

4.1.3. Components of accuracy

Schrader and Helmke (1987) proposed three components of TJA: (1) the *level component*, i. e., the correct estimation of the mean ability level of a class in contrast to a student's individual ability, (2) the *differentiation component*, i. e., the correct estimation of the variation of a trait within a class, and (3) the *comparison/rank component*, i. e., the assignment of the correct rank position of a student within his/her class. For the level component, findings indicate a rather accurate judgment (Spinath, 2005; Stang & Urhahne, 2016). However, some studies report a systematic overestimation (Bates & Nettelbeck, 2001; Madelaine & Wheldall, 2005). Regarding the differentiation component, results are heterogeneous as some studies reported an overestimation of the range in the class (Spinath, 2005; Südkamp, Möller & Pohlmann, 2008), some reported an underestimation (Helmke, Hosenfeld & Schrader, 2008), while others reported correct estimation (Stang & Urhahne, 2016). The rank order in the class, i. e., the comparison component, was found to be assigned accurately (Demaray & Elliott, 1998; Helmke et al., 2008; Spinath, 2005; Südkamp et al., 2008; Stang & Urhahne, 2016). Few studies have jointly considered all three accuracy components yet (Spinath, 2005; Südkamp et al., 2008; Stang & Urhahne, 2016). Spinath (2005) found that primary school TJA depended on both the accuracy component and the construct under scrutiny (e.g., intelligence, motivation), leading the author to conclude that "general diagnostic competency" does not exist. Additionally, Spinath (2005) found large individual differences in teachers' judgment accuracy.

4.1.4. Residuals as a measure of accuracy

Residuals refer to the distance between a score (e. g., TJs of cognitive ability) and its expected value (e. g., tested intelligence), representing the share of variance in a dependent variable that is not explained by independent variables (Colman, 2009).

Operationalizing accuracy as residuals originated in the field of teacher expectation research (e. g., Cooper, Findley & Good, 1982; Jussim & Haber, 2005). Holling and Preckel (2005) transferred it to judgment accuracy research by operationalizing the accuracy of (self-)judged intelligence through the residuals of predicting tested IQ from self-estimated intelligence. Sixty-eight percent of the sample were off in their estimates by a maximum of 8.51 IQ-points (ibid, p. 508). However, self-estimates of intellectual abilities were expected to be misjudged by 14.26 IQ-points, based on the meta-analytic findings of Mabe

and West (1982). Therefore, the authors found self-estimated intelligence to be more accurate than would be expected from previously reported correlations between actual and self-estimated cognitive ability.

Integrating the approaches described above, TJA was investigated (1) at a general level (i. e., as the correlation between TJs and IQ-test), (2) with respect to the three components of accuracy (i. e., level, differentiation, and comparison component) for TJs of their students' cognitive ability and (3) through residuals of predicting tested IQ from TJs of cognitive ability. As outlined above, results on TJA differ depending on their operationalization. To our knowledge, such an integrative approach has not been pursued before. Furthermore, the operationalization of accuracy through residuals is a relatively new approach (Baudson, 2011; Hinnant, O'Brien & Ghazarian, 2009; Holling & Preckel, 2005).

4.1.5. Influences on teacher judgment accuracy

TJs can be described as decisions made in a complex environment in which teachers gather unique, possibly contradicting information about each student (Heller, 2004). When judging or grading, teachers tend to rely on heuristics, thus reducing the complexity and potential information overload, but also bearing the risk of systematic judgment errors (Borko, Roberts & Shavelson, 2008). Common cognitive biases in TJ are the central tendency error, which means, that all judgments are within the average level regardless of possible manifestations towards the extreme, or the tendency towards extreme judgments, which means that judgments are either high or low, but not in between.

Südkamp et al. (2012) stated that TJA are influenced by several factors: (1) teacher characteristics, e. g., age and professional experience, (2) judgment characteristics, e. g. operationalization of judgment scales, (3) test characteristics, e. g., subject matter or test difficulty, and (4) student characteristics, e. g., motivation or knowledge. Typically studied teacher characteristics are age, gender and teaching experience, with differing results. Some researchers have reported significantly more accurate results for more experienced teachers (McElvany et al., 2009), while others did not (Praetorius, Karst, Dickhäuser & Lipowsky, 2011; Stang & Urhahne, 2016). Teachers' age and gender do not seem to affect TJA (Dicke, Lüdtke, Trautwein, Nagy & Nagy, 2012; Stang & Urhahne, 2016; Südkamp et al., 2012).

Several studies investigated how student characteristics affect TJs. The odds to be identified as intelligent or to be rated more intelligent than one's actual level of competence increase with socio-economic status (e. g., McBee, 2010; Valdez, 2013). In Germany, where the present study was carried out, academic success is highly dependent on socioeconomic factors (OECD, 2011; Prenzel, Sälzer, Klieme & Köller, 2013). German students from lower socio-economic backgrounds obtain lower grades than their higher-status classmates, even when controlling for objective ability (Maaz & Nagy, 2009). Kaiser et al. (2015), however, did not find systematic biases in judgment on the basis of the students' family background. This somewhat contradicting finding may be explained by its operationalization through number of books in the parents' household, instead of the parents' educational level or socio-economic status. Student gender is related to TJs as well, but current findings are inconclusive. As reported by Voyer and Voyer (2014), girls receive higher grades. Although gender differences in general cognitive ability are small to non-existent (e. g., Brunner, Krauss & Kunter, 2008), teachers have been found to ascribe higher intelligence to boys (Trautwein & Baeriswyl, 2007) or to girls (Mullola et al., 2012). Furthermore, ethnicity and migration background affect TJs. For the United States of America (USA), Tenebaum and Ruck (2007) showed in a set of four meta-analyses that European and Asian American students were favored over African American and Latino/a students. In comparison, Kristen (2006) found the impact of migration background on TJs to be rather negligible in Germany, but underlined the importance of German language skills.

4.2. The present study

We investigated accuracy of TJs of students' cognitive ability using different analytic approaches and factors biasing this judgment. Our research aims were fourfold:

1. We examined general accuracy of TJs of student cognitive ability. In line with previous findings, we expected a medium correlation (e. g., DeYoung, 2008; Machts et al., 2016).

2. Furthermore, we analyzed each of the three accuracy components (level, differentiation, and comparison). TJs of the level and comparison component were expected to be rather accurate, in accordance with most previous findings (Spinath, 2005; Südkamp et al., 2008; Stang & Urhahne, 2016). Findings regarding the differentiation component are somewhat inconclusive; therefore, no explicit assumption was made.

3. In addition, we analyzed the actual differences between tested intelligence and TJs of cognitive ability, i. e., the residuals from regressing TJs on IQ.

    a) As outlined above, systematic judgment errors often refer to the distribution of a trait. We therefore analyzed whether deviations of TJs from the objective measure were equally distributed over the ability spectrum or whether there was evidence for cognitive bias such as the central tendency error or the tendency towards extreme judgments.

    b) Analogous to Holling and Preckel (2005), we reported the standard error of estimation (SEE) from regressing TJs on IQ in IQ-points, because we were interested in the amount by which teachers were off in their judgment for the IQ-measure. Furthermore, we compared this SEE, to that, one may expect due to meta-analytic findings (Machts et al., 2016).

4. Finally, we related the residuals as a further measure of accuracy with students' gender, language background, age, and level of parental education, variables which are commonly examined as potential sources of bias in the context of TJ (e. g., Südkamp et al., 2012). We expected parental education to explain variance in TJs, as numerous studies have identified and replicated this relationship for Germany (e. g., Ehmke & Jude, 2010; Maaz & Nagy, 2009). Precisely, the higher the level of parental education, the higher the TJ of cognitive ability. No assumptions were made for students' gender, age, and language background, as findings are yet inconclusive (e. g., Baudson et al., 2014; Mullola et al., 2012; Trautwein & Baeriswyl, 2007).

4.3. Methods

4.3.1. Participants

Sample 1 (S1). The first sample comprised 679 primary school children in Grade 2 ($n = 354$, 49% female, 86% native German speakers) and Grade 3 ($n = 325$, 55% female, 86% native German speakers) along with their teachers ($N = 46$; Year 2/3: $n = 24/22$) in the federal state of Rhineland-Palatinate in Germany in June 2009. On average, students were 106 months old ($\sim$ 8.8 years; $SD = 8.27$ months). Two of the 46 teachers were male. On average, teachers were 42.59 years old ($SD = 12.43$; range 24–58), their professional experience ranging from 0–39 years ($M = 17.09$, $SD = 13.15$).

Sample 2 (S2). We assessed 2,079 students (Grade 1/2/3/4: $n = 336/428/394/921$) and their teachers ($N = 153$; Grade 1/2/3/4: $n = 27/31/29/66$) in primary schools in five German federal states (Hesse, Lower Saxony, Mecklenburg-Hither Pomerania, North Rhine-Westphalia, Rhineland-Palatinate) between September 2012 and January 2014. Forty-nine percent of the assessed students were female (range 46–53% per year) and 78.4% were native German speakers (range 76–79% per year). On average, students were 107 months old ($\sim$ 8.9 years; $SD = 15.03$ months). Twelve of the 153 teachers were male (8%), 19 did not indicate their gender (12%). On average, teachers were 42.02 years old ($SD = 11.02$; range 25–63) and had 15 years of professional experience ($SD = 11.25$, range 0–40).

4.3.2. Measures

Cognitive ability. Cognitive ability was assessed with the THINK 1–4 (for the piloting version used in S1 see Baudson and Preckel, 2013a, and for the final version used in S2 see Baudson et al., 2016). The 36-item THINK 1–4 assesses reasoning ability in the verbal, numerical and figural domain. The THINK 1–4 is applied as a paper-and-pencil test specifically designed for primary school students and group testing situations. The THINK 1–4 presents a reliable measure (Cronbach's Alpha $\alpha = .77$–$.82$) for assessing fluid intelligence. Criterion-related validity is supported by significant correlations with established intelligence tests, school-grades, self- and other-estimates.

Teacher questionnaire. The teacher rating scale of students' cognitive ability was developed based on established measures (see Baudson & Preckel, 2013a, for details). Teachers rated each child on three items mainly referring to fluid intelligence as in the students' potential. The items were (1) "is fast at identifying relations", (2) "understands new learning contents quickly', and (3) "can remember things upon first sight" (S1: 7-point Likert scale from 0 = *shows this characteristic less than all other students in class* to 6 = *shows this characteristic more than all other students in class*; S2: 6-point Likert scale from 1 = *fully applies to the student* to 6 = *does not apply to the student at all*). Furthermore, each teacher received one questionnaire asking for demographic data (teachers' age, gender, and professional experience).

Demographics and parent educational background. We collected demographic data (students' gender, mother tongue, parents' educational qualification) through parent questionnaires.

4.3.3. Procedure

Approval of the School Supervisory Boards and the data protection commissioners as well as parental consent were obtained. Students participated voluntarily and were rewarded with a toy (e. g., a pen), which they were not told in advance. Instructions to the intelligence test were administered verbally by psychologists or trained psychology majors and supported by overhead transparencies. Two weeks prior to the examination day, we gave one one-page questionnaire per child to the teachers. The filled teacher questionnaires were collected on the day the examination took place. We gathered demographic variables through parent questionnaires. The examination took between 90 (Year 1) and 45 minutes (Year 4).

4.4. Data Analysis

Beforehand, we analyzed descriptive statistics, reliability coefficients, and intercorrelations of measures. We calculated intra-class correlations (ICCs) of TJs because students from the same class are more alike than students attending different classes. We tested four levels of measurement invariance (MI) of TJ scales over school years: (a) configural invariance, i. e., invariance in the pattern of zero and non-zero loadings, (b) metric invariance, i. e., adding invariance in the size of factor loadings, (c) scalar invariance, i. e., adding invariance in the intercepts of the manifest variables, and (d) strict invariance, i. e., adding invariance of residual variances (Cheung & Rensvold, 2002; Widaman & Reise, 1997).

To address our first research question on general accuracy, we computed zero-order correlations between TJs of cognitive ability and the objective students' IQ. For our second research question referring to the components of TJA, analyses were performed as proposed by Spinath (2005). The level component (overall class ability level) resulted from the difference between the class-wise aggregated teacher's judgment and the students' actual IQ. For the differentiation component (range within a class), within-class variances of TJs and student intelligence were related (e.g., Feingold, 1992). A variance ratio of one means that variances in TJ and IQ are equal (i. e., accurate judgment). A ratio greater than one indicates an overestimation in TJ; a ratio smaller than one, an underestimation. For the comparison component (assignment of correct rank position within class), TJs of cognitive ability and IQ were ranked separately and analyzed through rank-order correlations. The continuous measure of IQ was therefore categorized into seven equally large

groups according to the seven-point Likert scale of the teacher rating questionnaire (for S2, the IQ measure was transformed into six groups of equal size).

To address research questions three and four, we standardized TJs within each class to control for the nested data structure, predicted students' IQ through TJs of cognitive ability, and saved the residuals. For our third research question, we analyzed the distribution of the residuals regarding teachers' over- (positive residuals) and underestimations (negative residuals) in relation to different levels of IQ. Furthermore, we converted the average distance between IQ and TJs of cognitive ability into IQ-points. Finally, to address our fourth research question referring to student variables affecting TJs, we correlated the residuals with student background variables (gender, age, mother tongue, and parents' educational background).

Measurement invariance tests and correlational analyses were conducted with M*plus*, Version 7.31 (Muthén & Muthén, 2015) using Restricted Maximum Likelihood (MLR) estimation, controlling for the nested data structure through M*plus*'s ANALYSIS TYPE= COMPLEX setting. For all other analyses, we used SPSS 21 (IBM Corp., 2012), using the class-wise standardized measure of TJs of cognitive ability.

4.5. Results

4.5.1. Preliminary analyses

Descriptive statistics, internal consistencies, and intra-class correlations. Descriptives of measures for both samples can be found in Table 4.1 (see page 26). All measures were sufficiently reliable. The ICCs indicated a notable influence of the classroom setting (S1: $\rho = .001$–004; S2: $\rho = .034$–$.091$). While the measure used class-based comparison standards in S1, general comparison standards were applied in S2. However, the influence of the classroom setting was even higher in S2. According to Barcikowski (1981), any ICC exceeding .01 should entail a consideration of the group level within the analyses. Although ICCs for TJs exceeded .01 in S2 only, we specified class membership as a cluster variable for both samples for reasons of consistency.

Measurement invariance. According to the cut-off of $\Delta\text{CFI} \leq -.01$ indicating non-invariance (Chen, 2007), TJs showed strict MI in both samples (see Table A.1 in Appendix A.1.1). Therefore, collapsing Years 2 and 3 (S1) and Years 1 through 4 (S2) as well as analyses at the manifest level were acceptable.

Table 4.1.: Descriptive statistics, reliability, and Intra-Class-Correlations of IQ and teacher judgments of cognitive ability

	N	*M (SD)*	*Min.*	*Max.*	Cronbach's α	ICC
Sample 1						
Grade 2						
IQ (THINK)	354	100.81 (14.12)	71	140	.80	
TJ_cognitive ability	353	3.47 (1.25)	0.00	6.00	.96	.004
Grade 3						
IQ (THINK)	325	100.87 (14.10)	72	135	.76	
TJ_cognitive ability	324	3.37 (1.37)	0.00	6.00	.97	.001
Grades 2 and 3						
IQ (THINK)	679	100.84 (14.10)	71	140		
TJ_cognitive ability	677	3.42 (1.31)	0.00	6.00	.96	.002
Sample 2						
Grade 1						
IQ (THINK)	333	102.92 (15.06)	72	150	.83	
TJ_cognitive ability	268	3.88 (1.05)	0.00	5.00	.96	.091
Grade 2						
IQ (THINK)	426	102.92 (15.06)	70	142	.85	
TJ_cognitive ability	349	3.50 (1.20)	0.00	5.00	.96	.070
Grade 3						
IQ (THINK)	390	99.90 (14.71)	71	142	.80	
TJ_cognitive ability	307	3.52 (1.14)	0.00	5.00	.96	.034
Grade 4						
IQ (THINK)	914	102.14 (14.56)	70	143	.82	
TJ_cognitive ability	817	3.55 (1.13)	0.00	5.00	.96	.078
Grades 1 through 4						
IQ (THINK)	2063	102.14 (14.56)	70	150		
TJ_cognitive ability	1741	3.59 (1.14)	0.00	5.00	.96	.082

Note. TJ_ = teacher judgment.

Zero-order correlations. Intercorrelations revealed similar patterns for S1 and S2 (see Table 4.2). TJ was significantly related to IQ ($r_{S1/S2} = .59/.53, ps < .001$) and parental education ($r_{S1/S2} = .28/.32, ps < .001$). As IQ and parental education were significantly related in both samples ($r_{S1/S2} = .23/.34, ps < .001$), we calculated partial correlations between parental education and TJ, controlling for IQ. In S1, TJ and parental education remained to correlate significantly with r = .12 (p > .01), the results were similar in S2 ($r = .15, p < .001$). The negative correlations between TJ and mother tongue ($r_{S1/S2} = -.16/-.10, ps < .001$) indicated higher judgments for German native speakers. Boys scored slightly higher in the piloting version of the THINK 1–4 ($r_{S1} = .09, S1$). Correlations between students' gender and other variables were very small, showing significance in S2 only, likely due to the larger sample size.

Table 4.2.: Intercorrelations of measures

		N_{S1}	N_{S2}	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
(1)	IQ (THINK)	679	2063		.03	−.05	.35***	−.14***	.53***
(2)	Gender	678	2063	.09*		.01	.05*	−.05*	.05*
(3)	Age	624	1928	−.06	−.03		−.15***	.00	−.13***
(4)	Parental education	631	1684	.22***	.03	−.13*		−.06*	.32***
(5)	Mother tongue	677	2060	−.24***	−.05	.06	−.07†		−.11***
(6)	TJ_cognitive ability	677	1741	.59***	.05	−.13**	.28***	−.16***	

Note. Gender: 1 = *female*, 2 = *male*. Mother tongue: 1 = *German*, 2 = *other*. TJ_ = teacher judgment. All correlations were computed with M*plus*, controlling for class membership through the TYPE=COMPLEX option. Correlations above the diagonal are those of *S*2.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

4.5.2. Research questions

General accuracy of teacher judgment of cognitive ability. TJs of cognitive ability and IQ correlated substantially in both samples ($r_{S1/S2} = .59/.53, ps < .001$).

Components of teacher judgment of cognitive ability. For the *level component*, the difference between judged and assessed cognitive ability was close to 0 (S1: $M = .03$, $SD = .38$, S2: $M = -.01, SD = .43$), indicating that teachers judged the level of cognitive ability in their classes rather accurately on average. However, the range revealed great variability between teachers (S1: −.93–.90; S2: −1.06–1.10).

For the *differentiation component*, the mean ratio was close to 1 (S1: $M = 1.04, SD = .25$; S2: $M = 1.01, SD = .32$), indicating that teachers judged the heterogeneity of their classes quite accurately. However, the range of the quotients revealed great variability between teachers (S1: 0.57–1.62; S2: 0.38–2.16; see Figure 4.1).

Regarding the *comparison component*, the median rank for TJs was $Md_{S1/S2} = 3.33/3.59$ and for IQ $Md_{S1/S2} = 4.00/3.00$. The rank order correlation revealed a large congruence between assigned and actual rank ($r_{S1/S2} = .64/.63, ps < .001$).

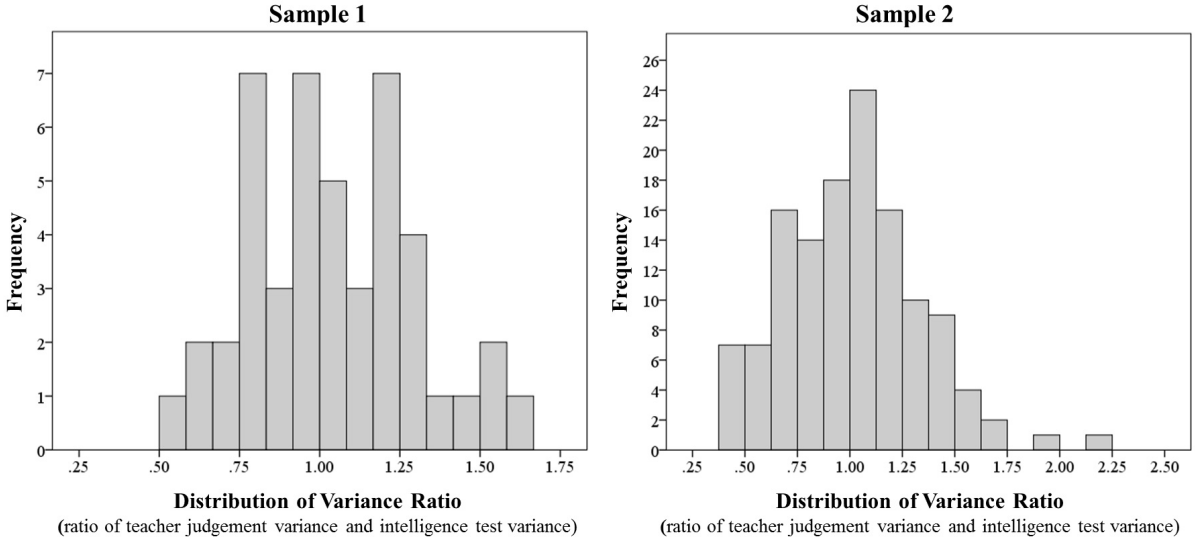


Figure 4.1.: Frequency distribution of the ratio of teacher judgment variance and intelligence test variance.

Pattern of teacher judgment accuracy. As described above, TJA was operationalized as residuals from regressing IQ on TJs. TJs of student cognitive ability predicted student IQ significantly in both samples ($\beta_{S1/S2} = .59/.53, ps < .001$; variance explained: 35% in S1, 28% in S2). When plotting the residuals over the IQ distribution (see Figure 4.2), teachers seemed to systematically overestimate the cognitive ability of low-IQ students and to systematically underestimate highly intelligent students in both samples. Comparing mean residuals over standard deviations of the IQ-distribution, univariate analyses of variance revealed higher TJA for students between IQ 85 and 115 than for students with below- and above-average IQs (S1: $F(3, 672) = 273.90, p < .001, \eta^2 = .550$; S2: $F(3, 1737) = 1030.48, p < .001; \eta^2 = .640$). We converted the SEE into IQ-points by taking the square root of the difference between 1 and the coefficient of determination, in this case, the squared correlation between TJs and IQ ($r = .50$), and multiplied it by 15, representing the non-restricted standard deviation of the standard IQ-scale (cf. Holling & Preckel, 2005). The SEE was 12.08 IQ-points in S1 and 12.46 IQ-points in

S1. Hence, about 68% of the teachers' judgments were off by up to 12.08 (S1) and 12.46 (S2) IQ-points, respectively. Using the non-restricted standard deviation of 15 IQ-points and the standard IQ-scale and the mean correlation of $r = .50$ between TJs of intelligence and tested IQ found in the meta-analysis of Machts et al. (2016), the SEE would have been 12.99 IQ-points.

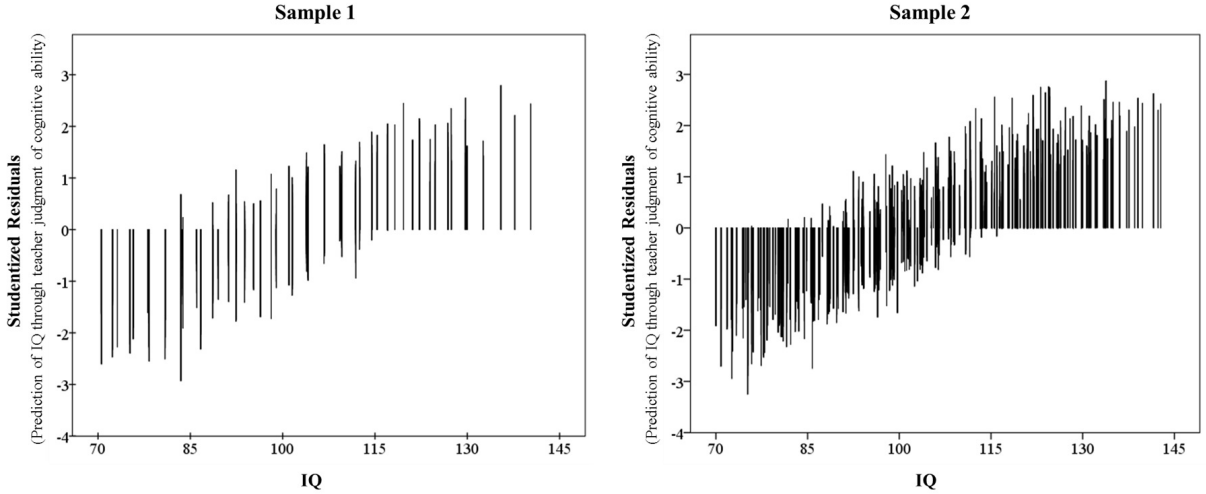


Figure 4.2.: Relationship between IQ level and residuals from predicting IQ through teacher ratings of cognitive ability.

Student characteristics related to teacher judgment accuracy. Correlations of residuals and student background variables showed a similar pattern in both samples. Student gender and age was not significantly related to the residuals from regressing IQ on TJs (for gender: S1: $r = .06, p = .118$; S2: $r = .01, p = .553$; for age: S1: $r = -.03$, $p = .399$; S2: $r = -.01, p = .653$). Teachers judged German native speakers' cognitive ability significantly higher ($r_{S1/S2} = -.18/-.11, ps < .001$). Parents' educational level was significantly related to the TJs and children of more educated parents were rated as more intelligent by their teachers (S1: $r = .11, p < .01$; S2: $r = .22, p < .001$).

4.6. Discussion

The present study investigated TJA of student cognitive ability using different methodological approaches, and possible sources of bias lowering TJA. We based our analyses on two large, independent samples of primary school students and their teachers. We analyzed TJA of cognitive ability from a broad perspective, combining different approaches

that covered individual as well as class level points of view and replicated our findings. Finally, our broad stance allowed us to compare different approaches to assess accuracy and to discuss their empirical as well as their practical implications.

4.6.1. Teacher judgment accuracy

General accuracy of TJ revealed correlations between TJs and the objective measure of large effect sizes (Cohen, 1988) in line with previous (meta-analytic) findings (e. g., DeYoung, 2008; Machts et al., 2016). Although the relation between TJs and IQ revealed large effects, only 35% (S1) and 28% (S2) of shared variance was found. Therefore, the large effect (according to Cohen, 1988) of a correlation might skew the picture of accuracy.

Regarding the different components, on average, teachers seemed to judge their class's ability level and heterogeneity as well as the rank order within their class rather accurately. This replicates earlier findings (Spinath, 2005; Stang & Urhahne, 2016). However, as in Spinath (2005), large inter-individual differences in teachers' ability to judge the heterogeneity and ability level of their classes were found. Generalizations of TJA are therefore questionable as they do not focus on inter-individual variability of teacher judgment ability. The comparison component revealed correlations of large effect size, indicating meaningful congruences between assigned and actual rank. Nonetheless, the squared correlation indicated only approximately forty percent of shared variance, i. e., about sixty percent of the variance was not explained, leaving room for the incremental influence of other variables.

Of note, while results of the meta-analysis by Südkamp et al. (2012) did not find a significant influence of norm-referenced versus peer-independent judgments, Machts et al. (2016) found TJA to be significantly lower in peer-independent judgments. While the measure used norm-referenced comparison standards in S1, peer-independent comparison standards were applied in S2. The results did not differ between the samples, in line with Südkamp et al. (2012) findings.

4.6.2. Pattern of teacher judgments' inaccuracy

Teachers misjudged their students' cognitive ability by approximately 12 IQ-points, in line with the expected value based on Machts et al. (2016) meta-analysis. That is, 68%

of the participating teachers misjudged their students' intellectual potential by slightly less than one standard deviation of the IQ-norm. Cognitive abilities of less able students were overestimated while more intelligent students were underestimated. The error of central tendency as a typical response bias may explain this pattern, since teachers tend to avoid extreme judgments and underestimate the students' heterogeneity (e. g., Spinath, 2005; Südkamp et al., 2008). Consequently, less able students may be demotivated when teachers overestimate their students' ability level and provide them with instructions that exceed their abilities. However, mild overestimations of a child's cognitive ability may lead to performance levels higher than what could be expected by potential alone (Baudson, 2011). Highly capable, but systematically underestimated students may not unfold their full potential due to cognitive underload and underchallenge, resulting in motivational losses and boredom (e. g., Heller, 2004; Preckel, 2008; Preckel, Götz & Frenzel, 2010). To conclude, while overestimations of less able students might have negative but also positive impacts on their development, underestimation of highly able students is likely to impact their development negatively.

4.6.3. Student-related sources of bias

Students' gender and age were unrelated to the residuals from regressing IQ on TJs. In line with previous findings, children of more educated parents were found to be overrated (e. g., Baudson et al., 2014; Maaz & Nagy, 2009; McBee, 2010). It can be argued that this overrating is justified, as parents who are highly educated themselves are expected to provide more support/resources for their children to develop their potential (e. g., Barbu et al., 2015). Furthermore, IQ is positively related to parental education, with 5 to 12% of shared variance; but even when controlling for IQ, the link between TJs and parental education remained. Thus, teacher's perception of the child's potential seems to be intertwined with its parents' educational background, which represents a violation of educational justice. Eventually, the relation of family educational background with TJs widens the gap between those who already possess material and educational capital and those who do not. Students whose mother tongue was not German were found to be systematically underestimated. As mother tongue was negatively related to parents' educational level, students from less educated parents were more likely to have a migration background. Therefore, students with a migration background might face a double disadvantage: first, because of their lower educational background, and second, because they are more likely to be lacking proficiency in German.

4.6.4. Different operationalization of teacher judgment accuracy

When comparing the different approaches to measuring accuracy, it may be concluded that depending on the research question a proper choice of operationalizing TJA is crucial, because results differ by method. When looking at general accuracy, the correlations between judged and tested IQ are of large effect size; yet detailed analyses show great inter-individual differences regarding the TJA of class level and students' heterogeneity. Furthermore, the regression-based approach showed that on average, teachers' ratings were off by 12 IQ-points. This means that the majority of participating teachers misjudged their students' intellectual potential by slightly less than one standard deviation of the IQ-norm. The results of the study show that the interpretations of large effect sizes as an indication for high accuracy might be a fallacy, as the variance explained, i. e., the degree of consistency, leaves much room for the impact of biasing variables.

The results point towards a distinct resemblance between general accuracy and the comparison component of TJA. For the comparison component, a continuous measure (e. g., IQ) is usually segmented into the same metric as categorical measures such as TJs. This reduces available information and also the scale of measurement artificially from continuous to ordinal. Because we do not see any notable differences between general accuracy and the comparison component, we would recommend not to reduce information by categorizing it and to use the general accuracy instead. Advantages of the regression-based approach of operationalizing accuracy by residuals are the normal distribution of such and the ability to analyze judgment deviations over the entire intelligence distribution. The normal distribution is the most common probability distribution and has several advantages, such as a mean of zero and a standard deviation of one, which makes it easy to work with mathematically (e. g., regarding transformations or inferences from the sample to the population).

Thus far, it remains unsettled how consistent the judgment and the objective measure should be in order to be considered accurate. Hypothetical perfect resemblance is possible, but in practice, one-hundred percent of shared variance between judged and measured characteristics is unlikely. In our study, we found correlations indicating large effects, which is rare in educational science. Nevertheless, they only indicate approximately one-third of shared variance of TJs and IQ, leaving much room for the influence of biases such as parental education. Overall, when examining TJA, the question of quality criteria remains unanswered, since correlations indicating large effects still lack a notable amount of consistency.

4.6.5. Implications for practitioners

From their meta-analysis, Machts et al. (2016) concluded that "expecting every teacher to have the competency to judge underlying intellectual potential may be unrealistic" (p. 100). Our findings revealed substantial inter-individual differences between TJA, pointing in the same direction. No matter whether teachers under- or overestimate the students' individual or the class ability level, it may be assumed that only an adequately challenging class setting may raise the students to fulfill their potential. Correct judgment, however, is a prerequisite for an adequately challenging class setting (e. g., McElvany et al., 2009). Furthermore, nomination behavior for special needs at either extreme of the intelligence distribution are dependent of accurate judgments. Teachers need better training in classroom assessment, as it is a necessary aspect of professionalization, but by far not standard in teacher education programs (Plake & Impara, 1997). For example, Hesse and Latzko (2009) proposed a 5-step procedure to self-assess diagnostic competencies: (1) selection of a student or test characteristic for which the teacher would assess his or her judgment ability, e.g., achievement motivation or task difficulty; (2) prediction of student behavior or performance; (3) assessment of student behavior or performance; (4) comparison of predicted and assessed behavior and performance; (5) analysis of the results of the comparison and search for possible reasons for discrepancies. Furthermore, they recommend teachers to work in professional groups to compare teaching and assessment methods.

4.7. Limitations and outlook

One limitation of the present study is that results are based on cross-sectional data, precluding any causal interpretation. The fact that the sample is German only prevents generalization across countries or cultures. For future research, further predictors could be considered. At the individual level, a further differentiation of the language background variable or standardized assessed levels of proficiency in German could prove insightful. Furthermore, parents' educational level represents a variable with limited explanatory power for psychological processes. Psychological variables, such as parental involvement, might help understand the processes behind (social) inequality. At the structural level, classroom descriptors like class size possibly impact TJA (e. g., Baudson et al., 2014).

4.8. Conclusion

In conclusion, the present study showed that in-depth analyses of the different components of TJA and patterns of divergence provide a much more differentiated picture than a focus on general judgment accuracy alone. Spinath (2005) concluded that the *one* diagnostic competence does not exist, as diagnostic competence depends on the constructs to be judged and on the accuracy component under scrutiny. We conclude that even when focusing on one construct, *one* diagnostic competence enabling teachers to judge their students' cognitive ability correctly does not exist, as results differ greatly by individual, component, and method. Considering the consequences of unfulfilled potential (as shown, e. g., for underachievers, Rimm, 2003), the investment in developing teachers' diagnostic abilities seems worthwhile not only for economic, but also for ethical reasons. Judgment procedures and errors should therefore be emphasized within any teacher education. In addition, mechanisms of judgment control (e. g., standardized classroom assessment and grading systems) might contribute to fairer assessments in our schools.

5. Keeping the gate: Components of primary school teachers' nominations for a gifted program in Germany

Abstract. Both researchers and practitioners agree that gifted education is necessary. In the process of selecting students for specific gifted programs, teachers play a crucial role. However, little is known yet about the dimensions underlying nomination decisions, and whether their respective importance differs by students' gender, considering that more boys than girls attend specific gifted programming. This study therefore examined predictors of teacher nominations and whether they differed for boys and girls. In a sample of 679 primary-school students in Grade 2 ($n = 354$) and Grade 3 ($n = 325$), we assessed cognitive ability, mother tongue, and parents' educational level. Teachers ($N = 46$) rated each student's cognitive ability, motivation, and creativity and gave a probability with which they would nominate each student for an enrichment program. Teacher nominations were explained by judgments of students' cognitive ability and creativity and parents' educational background. Teachers used differential approaches for girls and boys: Nomination of boys was predicted by judgments of cognitive ability; nomination of girls was additionally predicted by judgments of creativity and motivation, and by parents' educational background.

5.1. Theoretical background

Researchers and practitioners agree on the necessity of gifted education (e. g., Rogers, 2007). Accelerated and enriched curricula matching the gifted students' individual needs prevent cognitive underload, frustration, and boredom, among other issues (e. g., Baker, Bridger & Evans, 1988; Feldhusen & Moon, 1992; Heller, 2004; Preckel, 2008; Preckel et al.,

2010). One crucial element of any gifted education program is the successful identification of students whose needs and abilities match the offerings and requirements of the program (Mönks & Heller, 1994). Because identification processes using multiple methods and multiple informants are favored over single indicators, such as academic achievement or IQ-scores (Brown, Renzulli, Gubbins, Zhang, Siegle & Chen, 2005; Marland, 1972; Mönks & Heller, 1994). Furthermore, teachers are frequently asked to nominate students for gifted programs (Baudson, 2010). Teachers observe students in diverse settings varying in challenge over long periods, allowing them to gather unique information.

However, the role of teachers within the nomination process has been discussed controversially for decades of research on gifted education and gifted and talented programming (e. g., Gear, 1979). Most research has focused on the accuracy of teacher referrals in the identification of gifted children (e. g., Gear, 1979; Feldhusen et al., 1984; Gangé, 1994; Neber, 1994; Renzulli & Delcourt, 1986; Hanses & Rost, 1998). Less attention has been paid to the components underlying teacher nominations (e. g., Brown et al., 2005). Giftedness is a rather open construct whose conception is discussed controversially (Dai, Swanson & Cheng, 2011). When asked to identify gifted students, teachers seem to value skills related to academic performance such as cognitive ability over other student characteristics (Brown et al., 2005). Although high cognitive ability is the common denominator of most giftedness models (e. g., Sternberg, 1990), non-cognitive factors, such as self-concept, self-regulation or motivational characteristics, play an important role too (e. g., Gangé, 2004). Furthermore, findings show that besides cognitive and non-cognitive factors, sociodemographic characteristics of the students themselves are considered when teachers identify gifted students (e. g., parental socio-economic status/educational background: Valdez, 2013; gender: Petersen, 2013; ethnicity: Tenenbaum & Ruck, 2007).

The present study investigated German primary school teachers' nominations for a gifted program. In Germany, the system of gifted education is diverse and lacks a national strategy for education of gifted and talented students because education is in the hands of each federal state (Fischer & Müller, 2014). Giftedness and teaching gifted students is not a compulsory topic in most German university programs of teacher education. If, however, teachers are familiar with giftedness, one of the key working models is the Renzulli's Three-Ring Conception of Giftedness. The American Schoolwide Enrichment Model (SEM) by Renzulli and Reis (1985) is applied in Germany as well (e. g., Heller & Perleth, 2007). In the federal state of Rhineland-Palatinate, where the study was carried out, there is a comprehensive gifted program: the "Entdeckertag" (literally "discovery day"; Baudson, Wollschläger, Preckel & Vock, 2014), a weekly pull-out enrichment program aiming

at fostering highly intelligent, creative, motivated, and interested kindergarten and primary school students (Bildungsserver, 2016). The program is offered to two age groups (5–8 years and 7–10 years). Students explore and investigate a variety of topics (equivalent to the Type III-enrichment within the SEM model).

We examined how teacher judgments of different student characteristics and various student background variables impact the probability for a student to be nominated by his or her teacher for a gifted program[1]. By doing so, we aimed to shed light on components as well as possible sources of bias in teacher nominations for gifted programs. Furthermore, we compared findings for boys and girls, because boys are found to be 1.19 times more likely to be identified as gifted or participate in gifted programs (Petersen, 2013) and teachers have been found to nominate boys more frequently (e. g., Bianco, Harris, Garrison-Wade & Leech, 2011). To the best of our knowledge, our study is the first to examine components underlying nomination, and comparing them for boys and girls. We aim to contribute to our understanding of teacher nominations and of reasons for the imbalance between boys and girls in nomination-procedures for participation in gifted education.

5.1.1. Renzulli's Three-Ring Conception as a working model for components underlying teacher nominations

Teachers often think of gifted students as intellectually superior, highly analytically skilled, verbally advanced, motivated, and creative (Endepohls-Ulpe & Ruf, 2005; Rohrer, 1995; Copenhaver & McIntyre, 2012; García-Cepero & McCoach, 2009; but see also Baudson & Preckel, 2013c, and Preckel, Baudson, Krolak-Schwert & Glock, 2015, for negative stereotypes regarding the social-emotional adjustment of the gifted). This description of gifted students by teachers bears considerable resemblance to a popular model of giftedness in teacher education: Renzulli's Three-Ring Conception of Giftedness (Renzulli, 1978, 2005). According to this model, gifted behavior occurs at the intersection of above-average (intellectual) ability, creativity, and task commitment (motivation).

The positive influence of all three traits on academic achievement has been empirically supported in many studies. Intellectual ability and academic achievement are known to correlate positively and substantially (e. g., Deary, Strand, Smith & Fernandes, 2007).

[1]When referring to teacher judgment, we mean their judgment of students' characteristics, e. g., their cognitive ability. When referring to teacher nomination, we are talking about nominations for gifted programs.

Accordingly, intellectually gifted students tend to be academically successful (Roznowski, Hong & Reith, 2000). However, intellectual potential is no guarantee for success (Hanses & Rost, 1998; Lubinski & Humphreys, 1990). Task commitment predicts academic achievement positively (Elliot & McGregor, 2001; van Yperen, 2003). Creativity in terms of divergent thinking has been shown to be the second highest predictor of academic achievement after reasoning ability (Vock, Preckel & Holling, 2011).

Renzulli's model has been influential in teacher education about gifted students, especially with regard to the development of giftedness (e. g., Renzulli & Reis, 1985). Given that teachers' conception of gifted students aligns with Renzulli's model (e. g., Endepohls-Ulpe & Ruf, 2005; García-Cepero & McCoach, 2009) and that many gifted programs focus on these facets (e. g., Mönks & Heller, 1994), we aim to investigate how teachers' judgments of students' intellectual ability, task commitment (i. e., academic motivation and participation in class), and creativity relate to the likelihood of their nomination for a gifted program through their teachers.

5.1.2. Factors influencing teacher nominations for gifted programs

Since teacher nominations are typically the first step within a multi-step selection process (Heller, 2004), their role is crucial. On the one hand, teachers have the chance to identify and foster potentially gifted students, but on the other hand, they keep the gate by not identifying gifted students properly. In professional practice, teachers and/or hosts of gifted programs are often restricted to definitions of federal governments, state policy or education acts. For example, in the USA, 45 states (90%) include intelligence as a core concept in their definition of giftedness, 27 states (54%) of the states include creativity, as well, while only three (6%) include motivation (McClain & Pfeiffer, 2012). To our best knowledge, a comparable study for Germany remains to be done. However, the database "Giftedness" provided by the Karg Foundation in close cooperation with the respective German federal ministries (Karg Foundation, 2016), allows the summary that gifted and talented programs in German primary schools are either enrichment or acceleration programs for intellectually outstanding and/or highly achieving students. The federal states' policies do not state specific definitions, but rather stress that all children should be fostered to fulfill their potential (ibid.).

5.1.2.1. The role of cognitive ability

When it comes to identifying gifted students, teachers take cognitive abilities into account (e. g., Hernández-Torrano, Prieto, Ferrándiz, Bermejo & Sáinz, 2013; Neber, 1994; Siegle & Powell, 2004). Students' cognitive abilities strongly influence teacher's expectations regarding their students' performance (Baudson et al., 2014). However, high cognitive ability is usually equated with high academic achievement (Hany, 1997). Thus, instead of judging the students' potential, teachers tend to judge their performance or achievement (Baudson et al., 2014). If a highly able student is not achieving at a comparably high level, identification of being gifted is very unlikely (Hanses & Rost, 1998).

5.1.2.2. The role of task commitment/motivation

Clinkenbeard (2012) summarized contemporary motivation theories in the light of their importance for gifted students and gifted education. She concluded that whether considering expectancy-value theory, intrinsic and extrinsic motivation, goal theories or other concepts of motivation, all resemble explicit models and/or implicit theories about gifted/high-achieving students, and should therefore be incorporated into nomination procedures. Some studies have investigated the role of motivation within the identification process through teacher nomination. If students met a psychometric criterion for giftedness (e. g., IQ above 130, which means two standard deviations above the mean), but lacked motivation or showed a less serious attitude towards schooling, they were found to be less likely to be recommended for gifted education (e. g., Siegle & Powell, 2004). When looking at high-achieving students and their teachers' nomination for advanced courses, Barber and Tornay-Purta (2008) reported a significant advantage of those students reporting being intrinsically motivated and/or self-efficacious. Carman (2011) proposed an extension of traditional identification methods by personality questionnaires including measures of motivation, as they are at least in part capable of discriminating between gifted and non-gifted students; but at the same time, she stated the need of developing reliable and easily administrable instruments.

5.1.2.3. The role of creativity

Creativity has a rather ambiguous role in schoolhouse settings (Baudson & Preckel, 2013b). Although divergent thinking is positively related to academic achievement (Vock et al., 2011), teachers seem to associate high creativity with lower agreeableness and

lower conscientiousness (Karwowski, 2010). Prototypical creative students were also more likely to be chosen as least favorite students by their teachers (Westby & Dawson, 1995). The controversial findings on the relationship between creativity and achievement are not easily summarized. Some report a strong relationship between creativity, whereas others found weaker or negative connections (for a detailed review of the literature see Kaufman, Beghetto & Dilley, 2016). Since creativity plays a prominent role in theories of giftedness (see Kaufman, Plucker & Russell, 2012, for an overview), one may assume its importance within nomination procedures. But in fact, creativity tends to be overlooked (Pfeiffer, 2012), although instruments for the assessment of creativity exist (Kaufman et al., 2012).

5.1.2.4. Gender differences in teacher judgments

Although research findings suggest no significant difference in general cognitive ability between genders (e. g., Brunner et al., 2008), some studies report that teachers rated boys' cognitive ability higher than girls' (e. g., Trautwein & Baeriswyl, 2007), or vice versa (e. g., Mullola et al., 2012). As Südkamp et al. (2012) concluded in their meta-analysis of 75 studies on TJA, many more studies are needed that include teacher as well as student characteristics in the analyses of TJ.

Several, somewhat inconclusive findings reported gender differences in both, self-ratings and teacher perceptions of school-related motivational constructs. While boys score higher on global interest, girls score higher on global intrinsic motivation (Freudenthaler, Spinath & Neubauer, 2008; for domain specificity see, e. g., Spinath, Freudenthaler & Neubauer, 2010). Because motivational constructs are a conglomeration of various traits, findings regarding gender differences in TJs of students' motivation are diverse. Dicke et al. (2012) found students' mastery goals across various subjects to be higher for girls than for boys, but male students reported higher performance approach goals in Mathematics. Trautwein and Baeriswyl (2007) reported that teachers tend to perceive girls as more motivated. Mullola et al. (2012) found teachers to judge girls higher on motivation and persistence. Barber and Tornay-Purta (2008) reported an advantage of girls to be nominated for advanced courses in English when intrinsically motivated, and for math advanced courses when highly self-efficacious, whereas boys were more likely to be nominated in math when they were intrinsically motivated. Dicke et al. (2012) did not investigate gender differences in TJs of motivation itself, but considered students' gender as a moderator of TJA. They found teachers to be more accurate in judging boys in Mathematics (mas-

tery and performance-approach goals) as well as for mastery goals in the second foreign language (Dicke et al., 2012).

As we studied gender differences, we decided to choose two aspects of students' motivation for which gender differences have been shown. Regarding persistence, girls are rated higher by their teachers (Mullola et al., 2012). Furthermore, girls seem to be more engaged in the schoolhouse setting and better adjusted to educational demands (Legewie & DiPrete, 2012). Another gender difference has been shown for students' participation in class. Findings indicate that male classmates generally monopolize traditional face-to-face classroom discussion (e.ġ. Aukrust, 2008; Caspi, Chajut & Saporta, 2008). Aukrust (2008) investigated verbal participation in class starting in Grade one and found small but significant differences favoring boys already at this young age. Therefore, we included the teachers' judgment of their students' academic motivation focusing on persistence, effort and openness to new and possibly challenging tasks and their judgment of their students' participation in class/work attitude focused on oral participation and attitude of working in class. The above-mentioned findings suggest that teachers' perception of their students' motivation depends on the subject, the respective achievement goal orientation, and also on students' gender.

Findings on gender differences in creativity are rather inconclusive. Whereas female participants tend to outperform males on creativity tests, boys show more explorative behavior. Nevertheless, "a relative equality in creative ability" between girls and boys is likely (e. g., Baer & Kaufman, 2008, p. 76). Research on gender differences in teacher's judgment of creativity are scarce. Scott (1999) found teachers to judge girls as being generally more creative. However, whether gender differences are found or not seems to depend on absolute levels of creativity. Detailed analyses showed that in students of average creativity, only girls were perceived as more creative, whereas highly creative boys and girls were perceived equally creative by their teachers.

5.1.2.5. Gender differences in teacher nominations

Whereas girls generally receive higher grades (Voyer & Voyer, 2014), boys are more likely to be recommended for grade skipping or nominated for gifted programs (Bianco et al., 2011). Consequently, boys are 1.19 times more likely to be identified as gifted and/or included in gifted programs, as shown in a meta-analysis of 130 studies (Petersen, 2013). When teacher nominations or multiple criteria were used, no difference was found, but

when results of an IQ-test or a standardized achievement test were the sole criterion, boys were favored over girls (Petersen, 2013, p. 345).

5.1.2.6. Students' socio-demographic backgrounds and teacher nominations

Teachers' diagnostic competencies in general and especially for identifying gifted students in particular have been found to vary greatly between teachers (Siegle & Powell, 2004; Spinath, 2005). Variables like students' characteristics have been investigated as possible moderators.

Demographic and socio-cultural characteristics of the students have been found to bias TJs (e. g., Brown et al., 2005; Siegle & Powell, 2004). The most prominent biasing factor in Germany, where the present study was carried out, is the socio-economic and/or educational background of a student's parents. The odds to be identified as gifted or rated above one's actual level of competencies increase with socio-economic status (e. g., OECD, 2011; Prenzel et al., 2013; Valdez, 2013). Regarding ethnicity and migration background, meta-analytic findings suggest the highest advantage of Asian American students, followed by European American students, over African American and Latino/a students in the USA (Tenenbaum & Ruck, 2007). Compared to the USA, Germany's student population is rather homogenous. Only 20.3% of the German population have a migration background (Federal Statistical Office, 2015). Thus, ethnicity itself is rather less an issue than knowledge of the German language due to migration background.

All in all, when analyzing teacher judgments of intellectual ability, motivation, and creativity as components of teacher nominations in Germany, students' socio-economic status and migration background need to be taken into account. Although findings regarding gender differences are inconclusive, it is worth investigating their practical implications in terms of nominations for gifted programs because the ratio of participation in gifted programs favors boys. A gender-specific examination of components underlying teacher nominations may therefore point towards specific inequalities and thus contribute to a more nuanced discussion of gender disparities in gifted programming.

5.2. The present study

Nominating a student for gifted education is a decision made in a complex environment (Heller, 2004). When teachers observe students in diverse settings, they gather unique,

possibly contradicting information. When making decisions like nominating a student for a gifted program, teachers typically consider multiple factors, as well as heuristics, reducing the complexity and potential information overload (Borko et al., 2008). We examined which factors or components explain teacher nominations for a gifted program and whether these components differ for boys and girls. Our research aims were twofold:

1. We investigated which components of TJs and which student background variables explain the probability of nomination.

    a) First, we investigated the role of student background variables for teacher nominations. We investigated the impact of parents' educational background and students' gender and mother tongue on the probability of nomination. We expected to find a disadvantage of children of less educated parents or with a migration background. Regarding the influence of students' gender, findings are yet inconclusive; therefore, we did not frame any expectations.

    b) Second, we investigated the role of TJs for teacher nominations. Drawing on the components of Renzulli's model, we analyzed the teachers' judgment of their students' cognitive ability, motivation and participation in class, and creativity as predictors for nominating a student for a gifted program.

    c) Third, we investigated the role of student background variables and TJs for nominations simultaneously by integrating the above described analyses (1a and 1b) into one model to shed further light on the interaction between background variables and TJs when predicting nominations.

2. According to Petersen's meta-analytic findings (2013), on average, boys are 1.19 times more likely to be recommended for or participate in gifted programs. Explanations of this imbalance are rather speculative and the problem remains unsolved. For the two groups of boys and girls, we therefore compared the role of student background variables (i. e., parents' educational background, students' mother tongue, and IQ) and TJs (i. e., cognitive ability, motivation, participation in class, and creativity) in explaining teacher nominations.

5.3. Methods

5.3.1. Participants

Student sample. We assessed 679 primary school children in Grade 2 ($n = 354$) and 3 ($n = 325$) in the German federal state of Rhineland-Palatinate in June 2009. In total, 51.5% of our students were female (Grade 2 = 48.9%; Grade 3 = 54.5%) and 85.9% were native German speakers (Grade 2: 85.9%; Grade 3: 85.8%). On average, our students were 8.83 years old ($SD = .69$; Grade 2: $M = 8.32, SD = .46$; Grade 3: $M = 9.40, SD = .40$). We assessed 46 classrooms, with 9–24 students ($M = 14.76, SD = 3.97$) participating in each class (participation rate overall 68%). Attrition was due to missing parental approval or sickness on the day of the examination.

Teacher sample. Our teacher sample was comprised of the 46 main homeroom teachers of the 679 students. In Grade 2, we assessed 24 classes with different teachers, respectively, and in Grade 3, our sample consists of 22 classes with different teachers, respectively. Only two of the 46 teachers were male. On average, teachers were 42.59 years old ($SD = 12.43$; Grade 2: $M = 42.04, SD = 11.78$; Grade 3: $M = 43.18, SD = 13.36$). Teachers' professional experience ranged between 0 and 39 years ($M = 17.09, SD = 13.15$). In Grade 2, 14 of the 24 teachers had been their classes' homeroom teachers for Grades 1 and 2, while ten had taken over the class at the beginning of school year 2008/2009, i. e., almost 11 months prior to the examination. In Grade 3, the data is comparable. Seven of the 22 teachers had been their classes' homeroom teachers for all three years, two of the 22 teachers had taught their classes for two years, and 13 for one year.

5.3.2. Measures

Cognitive ability. Cognitive ability was measured with the T(H)INK (Baudson & Preckel, 2013a). The test assesses fluid intelligence (Gf) in the verbal, numerical, and figural domain through 12 items per domain, resulting in 36 items partly overlapping across grade levels. The test is described in more detail in (Baudson & Preckel, 2013a). The T(H)INK presents a reliable measure (Cronbach's Alpha $\alpha_{GRADE1/GRADE2} = .80/.76$) for assessing fluid intelligence. The T(H)INK-IQ is convergent and discriminant valid. The underlying hierarchical g-factor model shows an excellent fit. Criterion-related validity is supported by significant correlations with established intelligence tests, school-grades, as well as self- and other-estimates.

Teacher questionnaire. The rating scales used in the teacher questionnaire were taken or adapted from established measures (e. g., Gifted Rating Scales, School Form; Pfeiffer & Jarosewich, 2003; Child Behavior Check List 4–18/Teacher's Report Form; Achenbach & Rescorla, 2001; Social and Emotional School Experiences Survey/FEESS; Rauer & Schuck, 2003, 2004). Teachers rated each child on three items per scale. For the judgment of their students' cognitive ability, teachers were asked to rate the student's reasoning ability (e. g., "is fast at identifying relations" or "understands new learning contents quickly"). The judgment of their students' academic motivation focused on persistence, effort and openness to new and possibly challenging tasks (e. g., "tries to solve really difficult tasks" or "looks forward to new tasks and topics"), while the judgment of their students' participation in class/work attitude focused on oral participation and attitude of working in class (e. g., "works attentively and concentrated in class" or "enjoys participating in class"). For their judgment of creativity teachers were asked to judge their students' ability of divergent thinking and new insights within the learning process fostered by the student (e. g., "has many ideas" or "often finds unusual solutions to problems"). Furthermore, teachers judged some further characteristics, that are not reported here (e. g., social behavior). Altogether, the questionnaire was comprised 19 items (rated on 7-point Likert scales from 0 = *shows this characteristic less than all other students in class* to 6 = *shows this characteristic more than all other students in class*). Additionally, teachers were asked to rate the probability (0–100%) with which they would recommend each child for gifted programming ("How likely would you recommend this child for a gifted education program (0–100%)?").

Demographics and parent educational background. Demographic data like students' gender and mother tongue and parents' highest educational qualification were collected through parent questionnaires.

5.3.3. Procedure

We collected our data within a larger research project focusing on intelligence in elementary-school-children. The School Supervisory Board, the data protection commissioner of Rhine-land-Palatinate, and the students' parents approved of our study. Schools were recruited by phone, with the goal of obtaining a broad variety of school sizes, locations, and commuting areas. We tested all children in their classrooms. Test instructors were either experienced psychologists or trained psychology majors. Students participated voluntarily and we rewarded every student with a small toy. We handed out one questionnaire per child

to the teachers, prior to the examination day and collected the filled teacher questionnaires on the day the examination took place. Through parent questionnaires accompanying the parental approval form we collected demographic variables. Testing took between 75 minutes in Grade 2 and 45 minutes in Grade 3.

5.4. Data Analysis

For descriptive statistics, internal consistencies, and mean differences of teacher judgments for boys and girls, we used SPSS 21 (IBM Corp., 2012). All other data analyses were conducted with M*plus*, Version 7.31 (Muthén & Muthén, 2015) using MLR estimation. As students in our sample were nested within class-rooms, we ran all analyses (with the exceptions of the calculation of descriptive statistics) using students' class membership as a cluster variable within M*plus*'s ANALYSIS TYPE is COMPLEX setting.

Preliminarily, we calculated descriptive statistics, reliability of the measures, and correlations of measures, the latter revealing high intercorrelations between teacher ratings. Because students in the same class are more alike than students of different classes and a teachers' judgment are likely to be more similar to her own than to another teacher's judgment, we calculated intra-class correlations of all teacher-rated constructs. To ensure distinctiveness of the scales, we examined the factorial structure of the teacher questionnaire using confirmatory factor analysis (CFA) to test models with one to four factors against each other. MI of the measures used over grades, which is a prerequisite for collapsing data from students from different grades into one sample, was ascertained by applying a stepwise strategy to testing four levels of MI: (a) configural invariance (i. e., invariance in the pattern of zero and non-zero loadings), (b) metric invariance (i. e., additionally, invariance in the size of factor loadings), (c) scalar invariance (i. e., additionally, invariance in the intercepts of the manifest variables), and (d) strict invariance (i. e. additionally, invariance of residual variances; see Cheung & Rensvold, 2002; Widaman & Reise, 1997; Wu, Liu, Gadermann & Zumbo, 2010). We applied the same stepwise strategy for testing MI of the measures over gender, which is a prerequisite for comparing means and the conducted multi-group comparison.

Teachers rated the probability with which they would nominate a child for gifted programming on a scale from 0 to 100 percent. Thus, the rating may take on any value between 0 and 100. Our first research question addressed predictors of nomination probability.

To analyze the components influencing the probability of nomination, we conducted several regression analyses. The first model examined the influence of background variables (parents' educational background, students' gender, mother tongue) on nomination probability. The second model examined the influence of teacher judgments of cognitive ability, academic motivation, participation in class, and creativity on nomination probability. The third model took into account both background variables and teacher judgments.

Finally, our second research question addressed whether predictors of nomination probability differed by the student's gender. We used the third model of research question one considering both the background variables and the teacher judgments to compare boys and girls through a multi-group comparison. To compare the results of the multi-group analysis across gender we conducted gender invariance tests for all structural coefficients (βs).

For all regression analyses we partialed out the students' IQ, since students' cognitive ability has been found to act on teacher expectations regarding their students' performance highly (Baudson et al., 2014). However, we were interested in those differences in judgment, over and above actual IQ differences (see also Renzulli & Delcourt, 1986, who addressed the problem of teachers as "IQ guessers" more in detail).

5.5. Results

5.5.1. Preliminary analyses

Descriptive statistics, internal consistencies, and intra-class correlations. Descriptives are reported in Table 5.1. All measures were sufficiently internally consistent, except for the TJ of academic motivation. Taking together Grades 2 and 3, Cronbach's α for the TJ of academic motivation was $\alpha = .66$. Our measure consisted of only three items and covered different aspects of motivation (e. g., persistence and openness to new and possibly challenging tasks). Because teachers rated every student in his/her class, the use of a longer scale either covering more content to maximize the scale's validity or a longer one to maximize the scale's reliability was not possible. The ICC indicated a substantial influence of the classroom setting. According to Barcikowski (1981), any ICC with $\rho > .01$ should ensue a consideration of the group influence. The ICCs of TJs as well as the teacher nomination varied over grades. Hence, inter-individual differences between teachers in their judgments and nomination were revealed, i. e., teachers differed in their

judgments and probability of nomination. We controlled for this group influence by using class membership as a cluster variable within M*plus*'s ANALYSIS TYPE is COMPLEX setting.

Table 5.1.: Descriptive statistics, reliability, and Intra-Class-Correlations of IQ, teacher judgments, and probability of nomination

	N	*M (SD)*	*Min.*	*Max.*	Cronbach's α	ICC
Total sample						
IQ (THINK)	679	100.81 (14.10)	70.55	140.30		
TJ_cognitive ability	677	3.42 (1.31)	.00	6.00	.96	.002
TJ_academic motivation	658	3.24 (.96)	.00	6.00	.66	.057
TJ_creativity	663	3.26 (1.14)	.00	6.00	.91	.025
TJ_participation in class	671	3.51 (1.21)	.33	6.00	.85	.022
TJ_probability of nomination	679	32.85 (28.50)	.00	100.00		.243
Girls						
IQ (THINK)	350	99.67 (13.67)	70.55	137.71		
TJ_cognitive ability	349	3.36 (1.28)	.00	6.00	.97	
TJ_academic motivation	341	3.28 (.91)	.00	6.00	.68	
TJ_creativity	347	3.75 (1.15)	.67	6.00	.92	
TJ_participation in class	341	3.17 (1.10)	.00	6.00	.87	
TJ_probability of nomination	350	31.88 (28.13)	.00	100.00		
Boys						
IQ (THINK)	328	102.07 (14.49)	70.55	140.30		
TJ_cognitive ability	327	3.49 (1.33)	.00	6.00	.96	
TJ_academic motivation	321	3.20 (1.02)	.67	6.00	.67	
TJ_creativity	323	3.26 (1.23)	.33	6.00	.91	
TJ_participation in class	317	3.36 (1.18)	.33	6.00	.84	
TJ_probability of nomination	328	33.80 (28.91)	.00	100.00		

Note. TJ_ = teacher judgment. The Cronbach's Alphas (α) reported for the THINK are those reported by Baudson and Preckel (2013a). For separate descriptive statistics in Grade 2 and Grade 3 see Table A.2 in Appendix A.1.2.

Zero-order correlations. Intercorrelations of measures are depicted in Table 5.2 (see page 50). Except for student gender, all other variables were related to nomination probability. Teachers' nomination probability was most closely related to their judgment of students' cognitive ability ($r = .72$), followed by their judgment of creativity ($r = .65$)

and academic motivation ($r = .54$). Student gender was weakly related to TJs of creativity ($r = .08$), indicating that boys were rated as more creative, and TJs of participation in class ($r = -.20$), indicating that girls were perceived to participate more. Parents' educational background was significantly related to all TJs (rs = .17–.29) and the teacher nomination ($r = .28$).

Structural analyses of the teacher questionnaire. Because the scales of the teacher questionnaire were highly intercorrelated ($r = .55$–$.83$, all ps $< .001$; see Table 5.2, page 50), their distinctiveness was ascertained through CFA. We tested four models of one to four correlated factors. The model with separate but correlated factors for the TJs of cognitive ability, academic motivation, participation in class, and creativity fit the data best (see Table A.3 in Appendix A.1.2). Thus, the questionnaire assesses related but distinct facets of TJ.

Measurement invariance. According to Chen's (2007) cutoff point of $\Delta\text{CFI} \leq -.01$ indicating non-invariance, strict MI over grades could be shown for all TJ scales but academic motivation, where only partial strict MI was established (see Table A.4 in Appendix A.1.2), by fixing the negative residual variance of one item to zero in Grade 3, while in Grade 2, the residual variance was freely estimated. For further computations, Grades 2 and 3 could therefore be collapsed into one sample and analyses could be run with manifest variables (i. e., scale means). TJs of cognitive ability and creativity were strictly measurement invariant over the students' gender (see Table A.5 in Appendix A.1.2). We established partial strict MI for the TJs of participation in class, by fixing the negative residual variance of one item to zero for girls, while for boys the residual variance was freely estimated. For TJs of academic motivation only metric MI was established (ΔCFI between metric and scalar model = .06). Therefore, comparisons between TJs of academic motivation between girls and boys should be handled with caution.

Gender differences in teacher judgments and nominations. Teachers ascribed similar cognitive ability (girls: $M = 3.36, SD = 1.28$; boys: $M = 3.49, SD = 1.33$; $t(674) = -1.31, p = .19, d = -.10$) and academic motivation (girls: $M = 3.28, SD = .91$; boys: $M = 3.20, SD = 1.02; t(639) = 1.11, p = .27, d = .10$) to girls and boys. They rated their female students as participating significantly more in class than their male students (girls: $M = 3.75, SD = 1.15$; boys: $M = 3.20, SD = 1.23; t(668) = 5.36, p < .001$, $d = .41$). In contrast, boys were considered more creative (girls: $M = 3.17, SD = 1.10$; boys: $M = 3.36, SD = 1.18; t(656) = -2.11, p < .05; d = -.16$). Girls and boys were equally likely to be nominated for gifted programming (girls: $M = 31.88\%, SD = 28.13$; boys: $M = 33.80\%, SD = 28.91; t(676) = -.879, p = .38, d = -.10$).

Table 5.2.: Intercorrelations of measures

		N	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
(1)	IQ (THINK)	679								
(2)	Gender	678	.09*							
(3)	Parental education	631	.23***	.03						
(4)	Mother tongue	677	−.24***	−.05	−.07†					
(5)	TJ_cognitive ability	677	.59***	.05	.28***	−.16***				
(6)	TJ_creativity	658	.51***	.08*	.29***	−.15***	.83***			
(7)	TJ_academic motivation	633	.48***	.04	.19***	−.13**	.73***	.71***		
(8)	TJ_participation in class	671	.38***	−.20***	.17***	−.10*	.64***	.71***	.55***	
(9)	TJ_probability of nomination	679	.45***	.03	.26***	−.10**	.72***	.54***	.65***	.45***

Note. Gender: 1 = *female*, 2 = *male*. Mother tongue: 1 = *German*, 2 = *other*. TJ_ = teacher judgment. All correlations were computed with M*plus*, controlling for class membership through the TYPE=COMPLEX option.

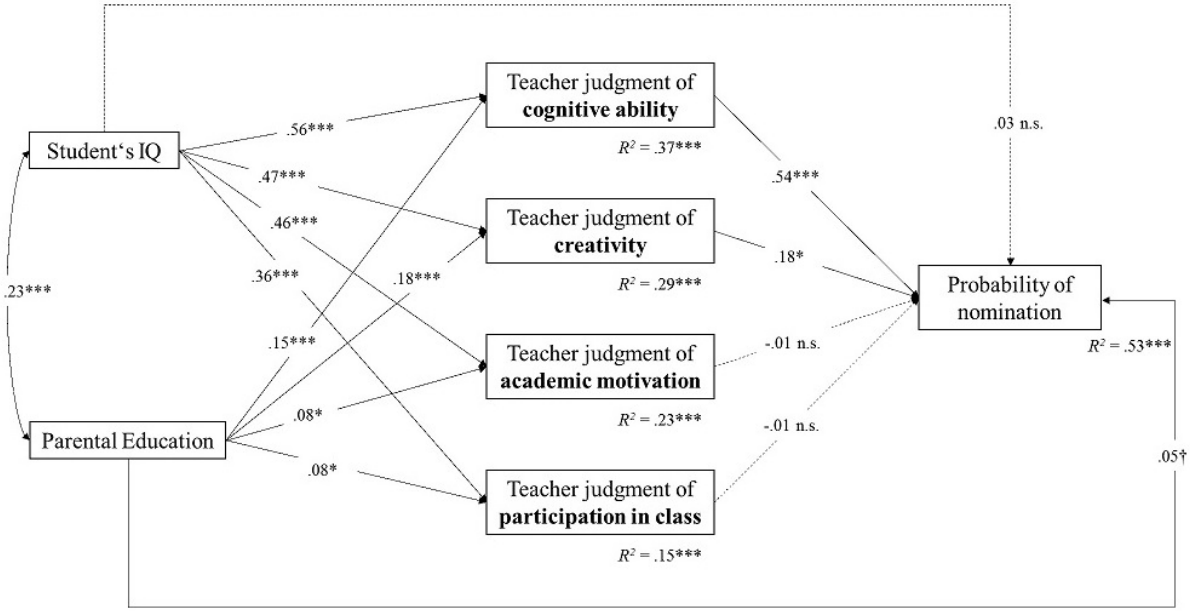
*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

5.5.2. Research questions

Explaining nomination probability through student background variables. The first regression model assessed the predictive power of student background variables. Probability of nomination was related to several background variables (see Table 5.2, page 50). The prediction of the nomination probability simultaneously through parental education, student gender, and mother tongue, and IQ showed a significant influence of parental education ($\beta = .16, SE = .04, p < .001$) over and above measured intelligence ($\beta = .43, SE = .04, p < .001$). The child's gender and mother tongue did not significantly predict nomination probability (gender: $\beta = -.01, SE = .04, p = .87$; mother tongue: $\beta = .02, SE = .03, p = .61$). The model explained 24% of the variance in nomination probability. Thus, IQ and parental education seemed to be essentially related to the probability of nomination, whereas neither students' gender nor mother tongue had incremental explanatory power.

Explaining nomination probability through teacher judgments. The second model included the TJs, showing a significant prediction through teacher judgments of cognitive ability ($\beta = .57, SE = .08, p < .001$) and creativity ($\beta = .19, SE = .07, p < .01$). In contrast, TJ of academic motivation ($\beta = -.01, SE = .07, p = .93$) and participation in class ($\beta = -.01, SE = .04, p = .724$) did not predict probability of nomination significantly. TJs explained 52% of the variance in nomination probability.

Explaining nomination probability through student background variables and teacher judgments. The third model included background variables as well as TJs as predictors. We found statistically significant influences of parental education and student intelligence on TJs that, in turn, influenced nomination probability. Results confirmed a significant influence of parental education and student IQ on all TJs (βs = .08–.56, ps $< .05 - .001$; Figure 5.1). Of the TJs, cognitive ability ratings ($\beta = .54, SE = .08$, $p < .001$) and creativity ratings ($\beta = .18, SE = .07, p < .05$) predicted nomination probability significantly. Additionally, a marginally significant direct effect of parental education remained over and above all TJs ($\beta = .05, SE = .03, p < .10$). This model explained 53% of the variance in nomination probability.

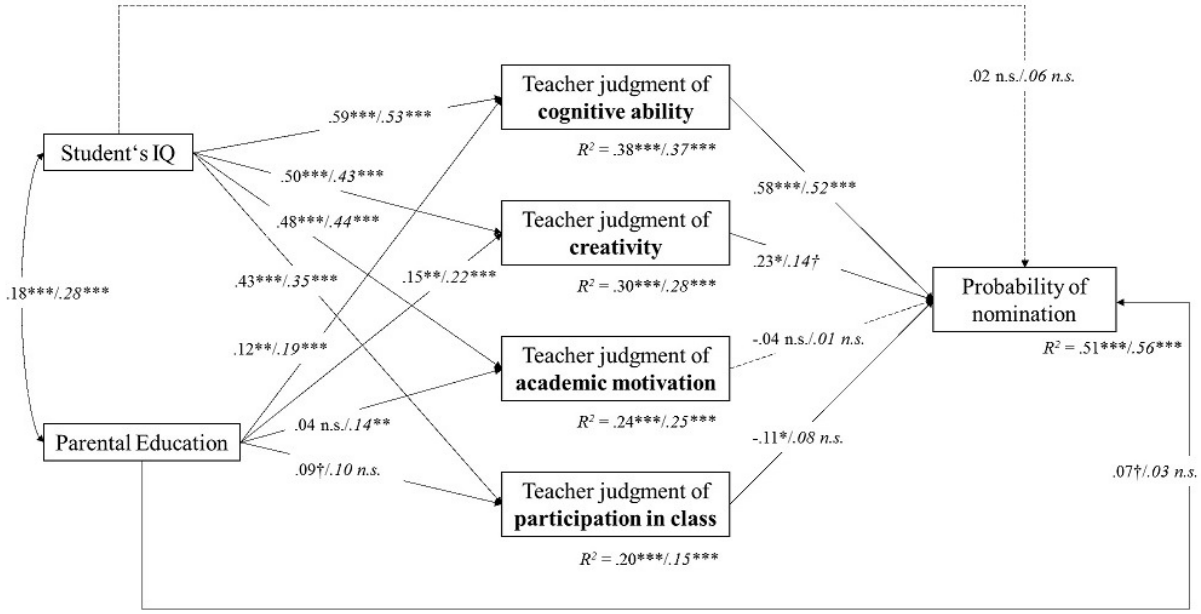


Note. Standardized model parameters are shown; teacher judgments as well as control variables are intercorrelated (see Table 5.2, page 50).

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Figure 5.1.: Regression model explaining the probability of nomination through teacher judgments and students' background variables.

Multiple-group comparison for boys and girls. A multi-group comparison (Figure 5.2) for the combined prediction model yielded a non-significant difference between boys and girls regarding the explained variance in nomination probability (girls: $R^2 = .51$; boys: $R^2 = .56; Z = .908; p > .05$), i. e., the predictor set explain the nomination probability equally well. Whereas for boys, TJs of cognitive ability predicted the probability almost on its own ($\beta = .52. SE = .09, p < .001$), for girls, TJs of creativity ($\beta = .23, SE = .11$, $p < .05$) and participation in class ($\beta = -.11, SE = .06, p < .05$) predicted nomination probability beyond TJs of cognitive ability ($\beta = .58, SE = .11, p < .001$). The marginally significant direct effect of parental education remained over and above all TJs for girls ($\beta = .07, SE = .04, p < .10$) but not for boys ($\beta = .03, SE = .05, p = .48$). The test of measurement invariance of all structural coefficients (see Table 5.3, page 59), revealed that the coefficients of nomination probability regressed on TJ of participation in class (for girls: $\beta = -.11, SE = .06, p < .05$; for boys: $\beta = .08, SE = .06, p = .145$) and those of TJs of academic motivation regressed on parental education (for girls: $\beta = .04, SE = .04$, $p = .290$; for boys: $\beta = .14, SE = .05, p < .01$) differed significantly. No other structural coefficients differed significantly from each other.



Note. Standardized model parameters are shown; teacher judgments as well as control variables are intercorrelated (see Table 5.2, page 50). Parameters in front of the dash are those of the girls, italic parameters following the dash are those of the boys.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Figure 5.2.: Regression model explaining the probability of nomination through teacher judgments and students' background variables by gender.

5.6. Discussion

By nominating students for gifted programs, teachers are important gate-keepers in gifted education. We examined student background variables (i. e., their parents' educational level, their mother tongue, and IQ and TJs of students in different domains (i. e. cognitive ability, motivation, and creativity) as predictors of teacher nominations for a gifted program. We further investigated whether these predictors differed for boys and girls.

What is new in our study is the investigations of teacher judgments underlying the nomination for gifted education and furthermore a gender-specific examination with regard to teachers' nomination behaviors, all while considering further well-known factors of influence (IQ: Baudson et al., 2014; parental socio-economic status/educational background: Prenzel et al., 2013; ethnicity: Tenenbaum & Ruck, 2007).

5.6.1. Predictors of nomination probability

We predicted nomination probability through relevant background variables and/or teacher judgments of cognitive ability, motivational variables, and creativity. When considering background variables only, student IQ and parental education significantly predicted the probability of being nominated for gifted programming, whereas student gender and mother tongue did not. The non-significance of gender as predictor is in line with Petersen's (2013) results. The result showed that only IQ measures and results of achievement tests favored boys over girls, but that when students were nominated through teachers, no gender differences were found (Petersen, 2013, p. 345). Furthermore, for our age group under study, no gender differences were reported (Petersen, 2013, p. 346). This finding could be considered as a hint that the imbalance of girls may not be due to differing nomination rates itself, but "losing" girls within the selection process, e. g., an advantage of boys is found when standardized achievement and intelligence tests are the core of the selection procedure (Petersen, 2013). Of note, one needs to keep in mind that in our study the nomination was rather hypothetical than of practical relevance. One might assume that practically relevant teacher nominations may be more restrictive and could be influenced through the program's detailed description (e. g., Mönks & Heller, 1994). Our results showed that teachers differed in their rated probability of nominating a student, i. e., some teachers were more likely to nominate students, while others were more restrictive. These inter-individual differences may reflect answering patterns like strict or mild answering or may be due to biases in teacher judgment. Common judgment errors are the central tendency error (i. e., all or most probability ratings are within the average range) or the tendency towards extreme judgments (i. e., probability ratings are categorized into very high and very low nomination probability whereas the center of the scale remains unused; e. g., Hesse & Latzko, 2009). Students' mother tongue, as an approximation of their migration background, did not predict nominations either. We assume that the mother tongue, i. e., migration background, was operationalized too broadly, as this group itself is rather heterogeneous. The importance of students' individual cognitive ability for teacher nominations is in line with previous findings (e. g., Spinath et al., 2010); thus, our findings support its central role in teacher nominations. The strong influence of parents' educational background replicates previous findings as well (e. g., Hernández-Torrano et al., 2013; McBee, 2010; Maaz & Nagy, 2009). Especially in Germany, the parental background influences the students' academic career profoundly (e. g., OECD, 2011; Prenzel et al., 2013). On the one hand, educational justice seems violated, since the child's potential and performance are intertwined with its parents' education and socio-economic status when judging/grading performance or uttering referrals. On the other

hand, when parents are highly educated themselves, we would expect them to be more involved and expectant regarding their offspring's education, providing needed support and resources (e. g., Barbu et al., 2015). One or the other, relating nominations to family educational background widens the gap between those who already have material and educational capital and those who do not.

With regard to TJs, we found that TJs of cognitive ability and creativity predicted nomination probability whereas TJs of academic motivation and participation in class did not. Considering Renzulli's Three-Ring Conception of giftedness (Renzulli, 1978, 2005), one may speculate whether teachers perceive giftedness as an overlap between all three characteristics or if they value one over the other. In line with previous findings and conceptions of gifted education, cognitive ability is valued most when nominating students (e. g., Baudson et al., 2014; Hernández-Torrano et al., 2013; Neber, 1994; Siegle & Powell, 2004). Creativity was found to predict the teacher nomination in a positive direction, meaning the more creative a child's teacher perceives him or her to be, the more likely he or she is to be nominated. This finding emphasizes the importance of considering creativity as a component of giftedness perception once more (e. g., Kaufman et al., 2012, 2016). Regarding motivation, one could speculate that teachers may not expect gifted children to be highly motivated in school because of academic underchallenge leading to boredom and frustration, among others (e. g., Baker et al., 1988; Feldhusen & Moon, 1992; Preckel, 2008; Preckel et al., 2010). That is, teachers might not consider motivation in school a reliable indicator of giftedness.

The combination of all predictors revealed that student IQ and parental education predicted all TJs whereas the probability of nomination was only predicted through TJs of cognitive ability and creativity. In addition, parental education had a marginally significant direct effect on the nomination. It seems that the background variables influence the TJ of specific students' characteristics more profoundly than the nomination itself. Besides providing an insight into what teachers take into account when nominating a student, these findings suggest that children of less educated parents suffer a double disadvantage: a direct one, because lower parental education is related to lower teacher ratings on all dimensions, and an indirect one, because lower cognitive ability and creativity judgments decrease the probability of these children being nominated. Of note, in all of these analyses we controlled for cognitive abilities of the children. That is, students of equal intelligence but lower family educational background are clearly disadvantaged when it comes to nominations for gifted programs.

5.6.2. Predictors of nomination probability for boys and girls

Girls and boys were equally likely to be nominated through teachers, replicating Petersen's (2013) meta-analytic findings. Regarding the components themselves, teachers judged cognitive ability and academic motivation similarly for both groups. The same judgments for boys and girls on academic motivation contradict some published research, stating boys to be perceived more capable whereas girls are perceived to be hardworking (Fennema, Peterson, Carpenter & Lubinski, 1990). However, female students' participation in class were judged higher than male students' participation, contradicting findings that male classmates generally monopolize traditional face-to-face classroom discussion (e. g., Aukrust, 2008; Caspi et al., 2008). Most of these studies investigated gender differences in older students (Kelly, 1988), except Aukrust (2008), who investigated verbal participation in class starting in Grade 1 and who also found small but significant differences favoring boys already at this young age. When looking at the items' wording in our study, participation in class comprised more than the traditional face-to-face interaction in our study. Participation also included conformity with rules, work organization, and work attitude (e. g., the students "works attentively and concentrated in class"). Therefore, findings may not be entirely comparable, since academic motivation and participation in class are related but distinct, and may vary with context across different studies.

In our study, teachers judged male students to be significantly more creative, while Scott (1999) reported higher ratings for girls than for boys. Since creativity is a rather broad construct, this finding raises the question of what teacher mean when judging it (Kaufman et al., 2016; Plucker, Beghetto & Dow, 2004). As high creativity seems to be linked to lower agreeableness and lower conscientiousness (Karwowski, 2010), one could assume that teachers mistake boys' prototypically more externalizing behaviors (e. g., Liu, 2004) for creativity. A current understanding of creativity emphasizes the role of creativity in schools, as the subjective experience, such that new insights and self-discovery during the learning process that are new and meaningful to the student may be considered creative activity as well (mini-c; e. g., Kaufman et al., 2016, p. 140ff. This understanding of creativity allows teachers to foster and value creativity within their classes and to choose teaching methods to transform (creative) potential into talent, but creativity, including its costs and benefits, has to be incorporated and advocated in teacher education (Pfeiffer, 2012).

Finally, we examined whether the underlying TJs on nomination probability were different for boys and girls. The most evident difference was that while TJ of cognitive ability predicted the nomination probability for both genders, an additional positive impact of

creativity judgments as well as a negative impact of participation in class was identified for girls. The difference between the relative importance of participation in class for the teachers' nomination was significant. Because our study did not focus on linguistic participation only, but included conformity with rules and work attitude as well, these findings might be explained by gender differences in teacher perceptions of gifted students. For example, teachers tend to attribute girls' success to hard work instead of innate ability (Fennema et al., 1990; Siegle & Reis, 1998). That is, the perception of a female student as being hard-working might be interpreted by the teacher as a compensation for lack of ability rather than an indicator of giftedness, while the perception of a male student as being hard working seems not be associated with teachers perceptions of his giftedness.

Parental education seemed to be influential rather on the TJs than on the nomination, although we found a direct influence over and above the TJs for girls and not for boys. Regarding TJ of academic motivation, the parental education seemed to be more influential for boys. When calling upon findings in the research area of parental involvement, parents of girls show more protective behavior (Jullien, 2006), whereas parents of boys seem to make contact with the school more often (Carter & Wojtkiewicz, 2000). Therefore, parents of girls and boys are involved, but the contact with the school may be of different nature. Nevertheless, the strong impact of the student's family background on educational success in Germany seem to be rather equally determining for boys and girls, instead of being limited to either one of them. These thoughts are rather speculative than empirically driven, for the question in how far gender differences in TJs influence the probability to be nominated for gifted programming has not been the focus of any but our study yet. Our results indicate that especially cognitive ability and creativity explain teachers' nomination, while girls might be disadvantaged when perceived as hard-working.

5.7. Limitations and outlook

Before concluding, some limitations of the present study must be pointed out. First, the results reported in this article are based on a cross-sectional design. Thus, findings cannot be interpreted causally. Second, the enrichment program prompting our research and thus our sample is restricted to one federal state in Germany (Rhineland-Palatinate); therefore, generalization of our findings requires further investigation. Third, teachers' nominations were hypothetical in nature.

For further research, it may be worthwhile to monitor the entire gifted identification process, starting with the nomination and ending with the actual participation and, eventually, success in the program, which could provide insights into selective dropout during selection and enrollment. Additionally, other background variables could be considered in further research. Possible influential variables on TJ might be, for example, parental involvement (Jullien, 2006), knowledge about giftedness itself and comparison standards teachers use (Rothenbusch, Zettler, Voss, Lösch & Trautwein, 2016). Furthermore, taking into account other motivational facets and/or investment traits, e. g., need for cognition, which leads individuals to engage in and enjoy effortful cognitive endeavors (Cacioppo & Petty, 1982), could provide insights into gifted identification as it has regarding students' attendance of special gifted classes (Meier, Vogl & Preckel, 2014). Additionally, the teachers themselves could add to our understanding of their role and influence on the nomination process. If some characteristics explaining the variability in, e. g., the number of students who are nominated as gifted, are identified, selection processes could be adjusted.

5.8. Conclusion

Taken together, the present study sheds light on the components, background variables, and their combination underlying nomination and on differences in the composition of these influential factors between girls and boys. We found two possible sources of bias within the nomination process. Parental education was strongly related to teacher nominations, especially for girls, and girls perceived as motivated, i. e., hardworking, by their teachers were less likely to be nominated. Based on previous and present findings, we draw the following conclusions: First, teacher education in general and for teaching gifted students in particular should cover sources of judgment bias. Especially, teachers need to be continuously reminded of the consequences of putting (too) much emphasis on students' parental background. And second, seeking for equal opportunities for boys and girls, as most gifted programs do, the relevance of our findings indicating differences within the underlying components should trigger further emphasis on gender differences in research and selection procedures, especially when considering the rather unexplained lower participation rates of girls in gifted education (Petersen, 2013). For girls (and boys, of course), working hard, seeking challenging opportunities, and being persistent in the pursuit of their goals should be understood as an additive asset in the development of excellence rather than a compensation for lack of innate ability.

Table 5.3.: Results of measurement invariance analyses of structural coefficients (β) of the multi-group comparison

		χ^2	*df*	Scaling correction factor	$\Delta\chi^2$ $(df), p$	CFI	RMSEA [90%CI]
Multi-group comparison model		0	0	1.000		1.000	.000 [.000 – .000]
Fixed Coefficient							
Criterion	*Predictor*						
TJ_probability of nomination	Parental education	.194	1	1.5490	.194 (1), .240	1.000	.000 [.000 – .110]
regressed on	IQ (THINK)	.423	1	.6019	.423 (1), .277	1.000	.000 [.000 – .124]
	TJ_cognitive ability	.337	1	1.0227	.337 (1), .743	1.000	.000 [.000 – .119]
	TJ_creativity	.570	1	1.4833	.57 (1), .240	1.000	.000 [.000 – .130]
	TJ_academic motivation	.289	1	.8136	.289 (1), .277	1.000	.000 [.000 – .117]
	TJ_participation in class	5.630	1	.9855	5.63 (1), .743	.997	.117 [.039 – .218]
TJ_cognitive ability	Parental education	2.546	1	.6946	2.546 (1), .240	.999	.068 [.000 – .176]
regressed on	IQ (THINK)	1.961	1	.6553	1.691 (1), .277	.999	.053 [.000 – .165]
TJ_creativity	Parental education	2.708	1	.6887	2.708 (1), .240	.999	.071 [.000 – .179]
regressed on	IQ (THINK)	.972	1	1.0392	.972 (1), .277	1.000	.000 [.000 – .143]
TJ_academic motivation	Parental education	4.093	1	.6121	4.093 (1), .240	.998	.096 [.014 – .199]
regressed on	IQ (THINK)	.022	1	1.4095	.022 (1), .277	1.000	.000 [.000 – .071]
TJ_participation in class	Parental education	.034	1	1.3420	.034 (1), .240	1.000	.000 [.000 – .080]
regressed on	IQ (THINK)	1.299	1	.8598	1.299 (1), .277	1.000	.030 [.000 – .151]

Note. To compare the structural coefficients between the groups, i. e., girls and boys, one model per coefficient was run, in which this path was set to be equal in both groups. If the adjusted model fitted significantly worse that multi-group comparison model in which all paths were freely estimated, the difference between the structural coefficients might be considered meaningful.

6. Teachers' secondary school track recommendations and parental enrollment decisions against students' social background

Abstract. Teachers' recommendation and parents' decision determine students' educational career within the tracked secondary school system in Germany. We investigated teachers' secondary school track recommendations and explained them through student's social background, school achievement, intelligence, as well as teacher judgments of their cognitive ability, motivation, and social behavior. Further, we investigated the correspondence between teacher recommendations and parent decisions and reasons for discrepancies. The sample comprised 597 fourth year primary school children attending 48 classrooms. We assessed students' intelligence and social background variables. Parents reported their final enrollment decision, student's school grades, and further social background variables. Homeroom-teachers (N = 48) rated each student's cognitive ability, motivation, and social behavior and made secondary school track recommendations. Teacher judgments of students' cognitive ability and motivation as well as the students' intelligence and school achievement explained secondary track recommendations while social background variables had no direct effect. However, social background variables were related to teacher judgments and school achievement. Parents largely followed teacher recommendations. Those who did not, seemed to consider their own social habitat more than the students' intelligence and achievement.

6.1. Theoretical background

International and national student assessment studies have shown the continuous importance of the students' social background for academic careers and achievement in Germany

(Maaz, Baumert & Trautwein, 2009; Jonkmann et al., 2010). The transition from primary to secondary school presents an influential hub within the academic career. At the end of primary school, teachers recommend a type of secondary school but in most federal states of Germany, the final decision where to enroll the student rests on the parents. Both, teacher recommendations[1], as well as parental enrollment decisions should primarily consider students' academic potential and achievement. However, in Germany the academic achievement and career are highly dependent on socioeconomic factors (e. g., OECD, 2011; Prenzel et al., 2013). Even when controlling for objectively assessed ability, German students from lower socio-economic backgrounds obtain lower grades than their higher-status classmates (e. g., Maaz & Nagy, 2009) and they are less likely to attend the highest school track (e. g., Maaz et al., 2009). That is, students' social background impacts the transition from primary and secondary school in favor of those students from educated, well situated households (e. g., Baumert, Maaz & Trautwein, 2009; Dumont, Maaz, Neumann & Becker, 2014; Klinge, 2016; Maaz et al., 2010).

The present study investigated teacher recommendations with a special focus on students' social background. We considered students' individual and family social background variables. We further considered students' academic potential and achievement as well as TJs on students' characteristics proven important for the transition from primary to secondary school (e. g., Anders et al., 2010). We aimed to replicate recent findings regarding teacher recommendations by a synopsis of multifaceted background variables using Raymond Boudon's idea of primary and secondary effects and their interaction on educational transitions (Boudon, 1974). We extended those findings by examining the degree of correspondence of teacher recommendations and parent decisions for final enrollment to contribute to our understanding of the impact of parents on tracking in secondary school.

6.1.1. Primary and secondary effects on educational transitions

In 1974, Raymond Boudon postulated that social inequality arises from the interplay of primary and secondary effects in educational transitions (Boudon, 1974). Primary effects are direct effects of achievement variables on educational transitions, e. g., the direct effect of the students' grades on track recommendations (Arnold, Bos, Richert & Stubbe, 2007). Secondary effects are direct as well as indirect effects of the students' social background

[1] While teacher recommendation always refers to the secondary track recommendation by the homeroom teacher, parental enrollment decision always refers to the parental final choice of secondary track enrollment.

on educational transitions independent of the level of academic performance, e. g., the influence of the parents' educational background (Neumann, Milek, Maaz & Gresch, 2010). Students' GPA, TJs of student characteristics, and achievement in standardized tests are typically investigated as primary effects, while secondary effects comprise parental socio-economic status (SES), their social and cultural background, or students' characteristics such as gender or migration background (etc.). Boudon's model has been the theoretical basis for many studies on social inequalities in education (e. g., Becker, 2016; Kloostermann, Ruiter, de Graf & Kraaykamp, 2009; Maaz et al., 2010; for an overview, see Glock, Krolak-Schwerdt, Klapproth & Böhmer, 2013; Dumont et al., 2014. Findings reveal that GPA is most closely related to secondary track recommendations, explaining 66 to 77% of variance (Arnold et al., 2007; Stubbe & Bos, 2008; Bos, Voss, Lankes, Schwippert, Thiel & Vaitlin, 2004). While achievement in standardized tests explains secondary track recommendations if considered solely, its increment over GPA seems to be negligible (Glock et al., 2013). TJs of students' motivation predicted track recommendations positively (e. g., Schneider, 2011; Stubbe & Bos, 2008), but not as strong as TJs of students' cognitive ability (Anders et al., 2010).

Regarding secondary effects, parents' SES is most influential in Germany (e. g., Bos et al., 2004; Maaz et al., 2010). If a student comes from a high SES family, it is three to four times more likely that he or she is recommended for a high track, even when controlling for ability and/or GPA (Glock et al., 2013). Cultural capital (i. e., number of books in the parents' household) has been found to be a significant predictor, too (e. g., Arnold et al., 2007; Stubbe & Bos, 2008) just as the highest parental graduation level, which was influential even after controlling for SES (e. g., Neumann et al., 2010). Regarding gender, some studies report recommendations in favor of girls (Arnold et al., 2007) while others report recommendations in favor of boys (Neugebauer, 2011). However, when controlling for academic achievement, girls and boys are equally likely to receive high-track recommendations (Glock et al., 2013). Students with a migration background are less likely to be recommended for a high track (e. g., Arnold et al., 2007; Gresch & Becker, 2010; Bos, Schwippert & Stubbe, 2007; Kristen, 2006). Again, when controlling for achieving, students' with and without migration background seem to be recommended for the high track with comparable likelihood (Glock et al., 2013).

Learning disorders have not been studied in the context of primary and secondary effects on secondary school tracking. However, teachers are confronted with students, who have special needs in learning how to read, write, and calculate, as well as staying attentive. The prevalence for learning disorders such as dyslexia, dyscalculia, and at-

tention deficit/hyperactivity disorder (ADHD) range between 2 and 8%, respectively (ADHD: 2–6%, Döpfner, Frölich & Lehmkuhl, 2013; dyscalculia: 4–7%, Jacobs & Petermann, 2007; dyslexia: 4–8% Warnke, Hemminger & Plume, 2004). Furthermore, findings regarding students with dyslexia and dyscalculia have shown that even with adequate intervention, the majority remains below average in their reading/writing or arithmetic ability after thirty months of treatment (Kohn, Wyschkon, Ballaschk, Ihle & Esser, 2015) and ADHD has been found to be a stable disorder as well (Döpfner et al., 2015). Therefore, learning disorders were integrated as potentially relevant student' background information within the teacher track recommendations.

Primary and secondary effects of the students' social background interact with each other (Boudon, 1974) and share common variability (e. g., students' grades and parental education). Therefore, primary and secondary effects should be considered together, in order to correctly estimate the effects and to understand their relative importance on secondary school recommendations. Maaz and Nagy (2009) were the first to investigate the relative importance of primary and secondary effects on test achievement, track recommendations and parental enrollment decisions. Results indicated that secondary and primary effects were equally important for tracking recommendations. While for achievement, primary effects were more important, for parental enrollment decisions, secondary effects had greater influence. Results indicated that social disparities did not derive from either primary or secondary effects, but rather originated from the interaction of students' social background and school grades, teacher recommendations, and parental enrollment decision. Neugebauer (2010) reported similar findings. Ditton (2013) focused on those students recommended for the lowest track, i. e., "Hauptschule". He found differences in achievement level, i. e., primary effects, to explain the recommendation as well as enrollment most and concluded that the impact of secondary effects might be more important for deciding between the high and medium than for the lowest track (ibid.). Meulemann and Relikowski (2016) found an increasing influence on primary effects and a decreasing influence of secondary effects, when comparing transitions from primary to secondary school between 1969 and 2007.

6.1.2. Teacher judgments and their relation to track recommendations

Anders et al. (2010) investigated the relation of TJs of students and their relation to track recommendations. They asked teachers to judge their students on a variety of factors

related to academic achievement and its development. Through explorative factor analyses they extracted three dimensions; (1) *talent and achievement* – including cognitive ability, psychological robustness, and academic skills, (2) *social abilities and behavior* – including the student's social behavior in class and impulse control, and (3) *motivation and learning virtue* – including learning/achievement motivation, effort, and discipline. TJs of talent and achievement were most predictive of teacher recommendations ($\beta = .75$), followed by TJs of motivation and learning virtue ($\beta = .20$). TJs of social abilities and behavior predicted recommendations negatively ($\beta = -.14$) which was explained by the authors as a suppression effect. Although regression weights decreased, this pattern of results remained when adding grades into the model.

Of note, TJs and academic achievement (i. e., grades) are impacted by secondary effects of students' social background. The most prominent predictors of TJs – when controlling for the objective level of the trait judged – are socio-cultural characteristics of students, their gender, and migration background (Gniewosz & Gräsel, 2011). Anders et al. (2010) found the SES to be related to all three TJ dimensions (*talent and achievement*: $r = .37$; *motivation and learning virtue*: r = .28; *social abilities and behavior*: $r = .19$). Students' gender was highly related to TJ of *social abilities and behavior* ($r = .32$) and *motivation and learning virtue* ($r = .23$), both judgments were in favor of girls. There was a small relation between the students' gender and the TJ of *talent and achievement* ($r = .07$), again favoring girls. Furthermore, all three dimensions of TJs were related to students' migration background ($r = .09 - .15$). Kristen (2006), however, found the impact of migration background on TJs to be rather negligible in Germany, but underlined the importance of German language skills for academic success.

6.1.3. Parental enrollment decisions

In most federal states of Germany, the final decision where to enroll the student rests on the parents. To explain parental enrollment decisions, two theoretical perspectives have been most prominent. The rational-choice perspective states that educational transitions are based on costs, benefits, and probabilities of success (e. g., Boudon, 2003). The other perspective focuses on the reproduction of the parents' own habitat (e. g., Bourdieu & Passeron, 2007). Both perspectives are related, because it can be assumed that highly educated parents value education more, and therefore the parents' own habitat is reproduced (Baumert, Maaz & Jonkmann, 2010). Ditton and Krüsken (2009) longitudinally investigated teacher recommendations and parental enrollment decisions. They found that more

than half of the recommendations as well as enrollment decisions were stable from Grade 2 onwards. Additionally, parents adjusted their aspirations toward the teacher recommendation. Differences in recommendation and enrollment especially concerned the high track "Gymnasium" and the lower track "Realschule" (ibid.). For a further understanding of parental tracking decisions, Klinge (2016) interviewed 25 parents and distinguished four groups of decision makers: (1) a group comprised of university graduates for whom the high track was self-evident; (2) a group comprised of university graduates, merchants, and craftsmen who obtained a critical distance to the schooling system but were educated themselves; (3) a group comprised of university graduates, merchants, and craftsmen who were not confident within the schooling system and aimed to secure the child's basic education; (4) a group of merchants and craftsmen who decided pragmatically (e. g., based on secondary school type availability). While parents of Group 1 mostly enrolled their children in the high track no matter what the teacher recommended, those of the other groups chose schools conform to the recommendation of the teacher or decided to "down"-track a child. Klinge (2016) concluded that differences in attitude towards schooling and the child's educational development were dependent on the parents own social status and habitat, resulting in group-specific enrollment behavior.

6.2. The present study

To give a context for the present study, we briefly describe the German school system. School systems of European countries differ greatly. Most federal states in Germany along with Austria and partly Hungary track their pupils after four years of joined learning, while other countries either track later (e. g., Liechtenstein after five, Luxemburg after six) or have a single structure educational system (e. g., Sweden; EACEA, 2016). Since school typically starts by the age of six, German pupils are tracked by the age ten. Students tend to remain in the secondary track in which they have started (Bellenberg, 2012; Lohmann & Groh-Samberg, 2010). Although students are able to switch between tracks in both directions, it is four times more likely to change to a lower track, than to a higher track (Berkemeyer et al., 2013). While the high track – the "Gymnasium" – allows students to pursue the A-level exams directly, and with it attend university, low tracks – such as the secondary schools with the tracks "Realschule" and "Hauptschule" – aim at qualifying students for the apprenticeships/vocational training. Bellenberg (2012) pointed out recent developments from a strict three-tier secondary school system toward a two-tier system, with the "Gymnasium" as the high track and the "Realschule" as the most

common secondary school track (see also Neumann, Maaz & Becker, 2013). Within this context we examined teacher recommendations and parental enrollment decisions against students' social background, academic achievement and potential, and TJs of cognitive ability, academic motivation, and social behavior. Our research aims were twofold:

1. In reference to Boudon's micro-theoretical model of education transitions (Boudon, 1974), we examined the relation of students' social background variables, their academic potential and achievement, and TJs of students' of cognitive ability, academic motivation, and social behavior to teacher recommendations (i. e., high track "Gymnasium" vs. lower tracks "Realschule" or "Hauptschule").

    a) Regarding primary effects, we examined the impact of students' GPA and TJs of their students' cognitive ability, academic motivation and social behavior, as well as the students' intelligence. We expected students' GPA and TJs of cognitive ability to be the most important positive predictors followed by TJs of motivation (e. g., Dumont et al., 2014; Anders et al., 2010). Furthermore, we assumed a higher probability of high-track recommendations for more intelligent students (e. g., Roos, Schöler, Zöller & Treutlein, 2011). Regarding TJs of social behavior, we did not make specific assumptions, because previous findings were somewhat inconclusive.

    b) Regarding secondary effects, we examined students' individual background variables (i. e., gender, age, migration background, learning disorders) and additionally students' family background variables (i. e., parents' educational level, cultural capital). We expected family background variables to predict teacher recommendations favoring those students with an educated family background.

    c) Applying Boudon's (1974) interactive model, we combined the primary effects model (1a) and the secondary effects model (1b) into one model (1c), aiming to further understand the relation of social background variables, TJs of their students' cognitive ability, academic motivation, social behavior, school achievement, and teacher recommendations.

2. We examined in how far the teacher recommendation corresponds with the parents' final enrollment decision. We investigated group differences regarding students' individual and family background variables between those parents who followed teacher recommendations and those who did not trying to explain divergent parental decisions by student and parent variables. In accordance with the findings by Klinge

(2016), we expected that divergent decisions could be explained by the parents own social status and habitat.

6.3. Methods

6.3.1. Participants

Student sample. We collected our data in a larger research project focusing on intelligence assessment in elementary school children. We assessed 597 children in Grade 4 in 48 classes in the federal states of Rhineland-Palatinate and North-Rhine-Westphalia in Germany between March 2013 and February 2014. Table 6.2 (see page 78) gives a detailed description of the sample. Of the students, 49.9% were female and 79.1% were German native speakers. On average, the students were 10.09 years old ($SD = .49$). About half of the sample (49.7%) were recommended for the high track by their teachers; parents of 48.7% of the students decided for the high track. Parents of most students (87%) chose the track recommended by teachers; 44% decided for the high track and 43% for a lower track. Fourty-two students (7%) enrolled in the high track without the matching teacher recommendation, while 36 (6%) students enrolled in a track lower than their teachers' recommendation.

Teacher sample. The teacher sample comprised 48 homeroom teachers (43 female). On average, teachers were 40.16 years ($SD = 11.53$, range 26–63) old and had 12.94 years of professional experience ($SD = 11.36$, range 0–40).

6.3.2. Measures

Cognitive ability. Cognitive ability was assessed with the THINK 1–4 (Baudson et al., 2016). The THINK 1–4 assesses Gf in the verbal, numerical and figural domain. It is specifically designed for primary school children and group testing situations. The THINK 1–4 is applied as a paper-and-pencil test and due to its generous time limits, classified as a power test. THINK 1–4 presents a reliable measure (Cronbach's Alpha $\alpha = .77 - .82$; Split-half retest-reliabilities after 16 months between $r_{tt} = .71 - .77$; ibid.) for assessing fluid intelligence in group-settings. THINK 1–4 results correlated significantly with established intelligence tests ($r = .55 - .78$; all ps $< .01$; ibid.).

Teacher questionnaire. Teachers filled out two types of questionnaires, one assessing demographic variables such as gender, age and professional experience and the other focusing on judgments of student characteristics. Items of the TJ scales were taken or adapted from established measures (see Baudson & Preckel, 2013a, for details). Teachers judged each child on 15 items (6-point Likert scale from 1 = *fully applies to the student* to 6 = *does not apply to the student at all*). Teachers judged their students' cognitive ability (e. g., "understands new contents quickly") on six items, covering reasoning ability and knowledge acquisition. The judgment of their students' academic motivation on six items focused on persistence, effort and openness to new and possibly challenging tasks (e. g., "tries to solve really difficult tasks"), as well as their students' oral participation and attitude of working (e. g., "works attentively and concentrated in class"). Teachers judged their students' social behavior on three items, covering the students' social attitude towards other children and his/her teacher(s) (e. g., "gets along well with other children"). Additionally, we asked teachers to give a secondary track recommendation for each child; categories were "Gymnasium" (highest track), "Realschule/Hauptschule" (lower tracks) and integrated secondary school, "Gesamtschulen", with the track specification.

Parent questionnaire. We collected student background variables (gender, age, migration background, learning disorders), family background variables (parents' educational level, cultural capital), indicators for previous educational history (GPA, grade retention or acceleration), and the parents' final choice of secondary school track through parent questionnaires. For learning disorders, we included dyslexia, dyscalculia, ADHD and any combination of those, if professionally diagnosed. Cultural capital was operationalized through the (estimated) number of children's books possessed by the student's family. Moreover, parents were asked to report the students' grades at the time of data collection (i. e., from the current report card).

6.3.3. Procedure

The School Supervisory Board and the data protection commissioner of Rhineland-Palatinate approved the study. The students' parents approved their child's participation. We recruited schools by phone, with the goal of obtaining a broad variety of school sizes, locations, and commuting areas. We tested children in their classrooms with either experienced psychologists or trained psychology majors as test instructors. Students participated voluntarily. Two weeks prior to the examination day, we handed out one questionnaire per child to the teachers. We collected the filled-out teacher questionnaires on the day

of the examination. Through parent questionnaires accompanying the parental approval form, we collected demographic variables. The examination lasted approximately 45 minutes. Assessment was completely anonymous. The students' tests, teacher, and parent questionnaires had corresponding codes to guarantee correct assignments between the questionnaires.

6.4. Data Analysis

For descriptive statistics, internal consistencies, and partial correlations, we used SPSS 21. All other data analyses were conducted with M*plus*, Version 7.31 (Muthén & Muthén, 2015) using restricted MLR estimation. Preliminarily, we calculated descriptive statistics, reliability of the measures, correlations of measures, and intra-class correlations of all teacher rated constructs. As students in our sample were nested within classrooms, we ran all analyses (except the calculation of descriptive statistics) using students' class membership as a cluster variable within M*plus*'s ANALYSIS TYPE is COMPLEX setting. For correlations run with SPSS 21 (IBM Corp., 2012), we used class-wise standardized variables.

Our first research questions addressed primary and secondary effects of social background variables on teacher recommendations. To analyze the variables influencing teacher recommendations, we conducted two logistic regression analyses using Montecarlo integration, as we had missing data on those variables collected through parents' questionnaires (covariance coverage between 84.2% and 100%; lowest for students' age). The first model examined the impact of primary effects on teacher recommendations comprising students' GPA, IQ, and TJs of cognitive ability, academic motivation, and social behavior. The second model examined secondary effects of individual and family background variables (i. e., gender, age, migration background, learning disorders, parents' educational level, and cultural capital) on teacher recommendations. Finally, we conducted a logistic path analysis considering both, primary as direct and secondary effects as direct and indirect predictors of teacher recommendations. In this model, student and family background variables as well as IQ predicted TJs and GPA, which in turn predicted teacher recommendations. Background variables as well as IQ also directly predicted teacher recommendations.

For our second research question, we analyzed the relation of teacher recommendations and parent decisions. We separated the sample into four categories: (1) students with teacher recommendation and parental decision for the high track; (2) students with teacher

recommendation and parental decision for the lower tracks; (3) students with teacher recommendation for the high track and parental decision for lower tracks, i. e., "Down-Tracker"; (4) students with teacher recommendation for lower tracks and parental decision for the high track, i. e., "Up-Tracker". We compared the groups regarding their social background applying analysis of variance (ANOVA)s with Tukey's honest significance difference (Tukey's HSD) mean separation test for post-hoc analyses (Winer, 1971) for interval scaled variables and Kruskal-Wallis H-Tests for ordinal scaled variables.

6.5. Results

6.5.1. Preliminary analyses

Descriptive statistics, internal consistencies, and intra-class correlations. Descriptive statistics are reported in Table 6.1. All measures were almost normally distributed and showed sufficient reliability (one exception: TJs of social behavior). The ICC indicated a substantial influence of the classroom setting on all TJs. According to Barcikowski (1981), any ICC with $\rho > .01$ should lead to a consideration of the group influence within the analyses. We controlled for this group influence by using class membership as a cluster variable within M*plus*'s analysis "type is complex" setting.

Table 6.1.: Descriptive statistics, reliability and Intra-Class-Correlations of IQ-measure and teacher judgments

	N	***M (SD)***	***Min.***	***Max.***	**Cronbach's α**	**ICC**
IQ (THINK)	597	103.1 (14.75)	63.32	138.50	.80	.105
GPA	576	4.8 (.73)	2.50	6.00	.84	.023
TJ_cognitive ability	597	3.57 (1.07)	.00	5.00	.96	.061
TJ_academic motivation	597	3.55 (1.02)	.00	5.00	.90	.106
TJ_social behavior	597	4.07 (.82)	.00	5.00	.75	.235
TJ_track recommendation	597	49.7% high track recommendation				.006
Parental decision	597	48.7% high track enrollment				.011

Note. TJ_ = teacher judgment. GPA refers to the mean grade in German, Math and General Studies.

Correlational analyses. Correlations of measures are depicted in Table 6.3 (see page 79). Teacher recommendation was significantly related to all other variables, even after controlling for students' IQ and GPA. The highest relationships were found between teachers' track recommendation and their judgments of cognitive ability ($r = .73$), academic

motivation ($r = .63$), and the students' GPA ($r = .72$). The high relationship between track recommendation and TJs remained for TJs of cognitive ability and of academic motivation, when controlling for IQ (TJ of cognitive ability: $r = .46$; TJ of academic motivation: $r = .43$), GPA (TJ of cognitive ability: r = .27; TJ of academic motivation: $r = .25$), or both (TJ of cognitive ability: $r = .21$; TJ of academic motivation: $r = .23$).

6.5.2. Research questions

Primary effects on teacher recommendation. The logistic regression model explaining teacher recommendations by TJs, students' GPA and intelligence (see Table 6.4, page 80) showed a significant relation with GPA ($\beta = .42, SE = .07, p < .001$), intelligence ($\beta = .12, SE = .04, p < .01$), TJs of cognitive ability ($\beta = .31, SE = .07, p < .001$), and TJs of academic motivation ($\beta = .23, SE = .07, p < .01$). TJs of social behavior ($\beta = -.07, SE = .05, p = .14$) did not predict teacher recommendations. Overall, the model explained 68% of the variance in teacher recommendations.

Secondary effects on teacher recommendations. The explanation of teacher recommendations through student and family background variables showed significant effects of both (see Table 6.4, page 80). Girls ($\beta = -.11, SE = .05, p < .01$), German native speakers ($\beta = .13, SE = .04, p < .01$), and younger students ($\beta = -.18, SE = .05, p < .001$) were more likely to be recommended for the high track, while students with learning disorders ($\beta = -.32, SE = .05, p < .001$) were less likely recommended. Parents' educational background predicted teacher recommendation positively ($\beta = .31, SE = .05, p < .001$); those who possessed more children books where more likely to obtain a high track recommendation ($\beta = .07, SE = .04, p < .05$). Overall, the model explained 34% of the variance in teacher recommendations.

Primary and secondary effects on teacher recommendations. Results of the logistic path analysis considering both, primary effects as direct predictors and secondary effects as direct and indirect predictors of teacher recommendations are reported in Table 6.5 (see page 81). The students' IQ and age were related to all TJs (IQ: βs = .18–.53, ps < .001; Age: βs = −.12– −.15, ps < .001). Girls received higher grades ($\beta = -.06, SE = .03, p < .05$) and were judged to be more motivated ($\beta = -.19$, $SE = .04, p < .001$) and socially competent ($\beta = -.18, SE = .04, p < .001$). Learning disorders were negatively related to students' GPA ($\beta = -.21, SE = .03, p < .001$), TJs of cognitive ability ($\beta = -.14, SE = .04, p < .001$), and TJs of academic motivation ($\beta = -.17, SE = .05, p < .001$). Students' mother tongue only predicted TJs of

cognitive ability favoring German native speakers ($\beta = .06, SE = .03, p < .05$). Family background was significantly related to GPA and TJs of cognitive ability and academic motivation with higher effects for parental education (GPA: $\beta = .15, SE = .04$, $p < .001$; TJs of cognitive ability: $\beta = .16, SE = .03, p < .001$; TJs of academic motivation: $\beta = .12, SE = .04, p < .01$) than for the number of children's books (GPA: $\beta = .08, SE = .03, p < .01$; TJs of cognitive ability: $\beta = .04, SE = .02, p < .10$; TJs of academic motivation: $\beta = .05, SE = .03, p < .10$).

All TJs predicted teacher recommendations significantly: cognitive ability ($\beta = .31$, $SE = .07, p < .01$), academic motivation ($\beta = .22, SE = .07, p < .01$), and social behavior ($\beta = -.08, SE = .05, p < .10$). Students' GPA showed the strongest relation with teacher recommendations ($\beta = .40, SE = .06, p < .001$). Additionally, a significant direct effect of students' IQ remained over and above the impact of TJs and GPA ($\beta = .11, SE = .05, p < .01$). Overall, the model explained 71% of the variance of teacher recommendations.

Parental enrollment decision. Most students (86.9%) were enrolled in the track recommended by teachers and parental decisions and teacher recommendations were highly related ($r = .74, p < .001$). Parents' decision correlated significantly with parental educational background ($r = .40, p < .001$), also when controlling for teacher recommendations ($r = .22, p < .001$). Regarding the final parental enrollment decision, we compared four groups with each other as described in the data analysis section. The description of each group is displayed in Table 6.2 (see page 78). Students in the four groups differed most regarding their IQ, $F(3, 593) = 99.20, p < .001$, and GPA, $F(3, 572) = 166.03, p < .001$. Post-hoc analysis revealed that independent of final enrollment, those students with a high track recommendation did not differ significantly in their intelligence (Group 1: M $[SD]$ = 111.96 [11.83], Group 3: $M[SD]$ = 107.71[12.30]), while they were significantly more intelligent than those with a low track recommendation (Group 2: $M[SD]$ = 94.01[12.61], Group 4: $M[SD]$ = 98.75[9.70]). Comparing Groups 1 and 4, which are comprised of students whose parents decided for the high track, those with a teacher recommendation for a high track scored 13 IQ-points higher than those with a lower-track recommendation. Comparing Groups 2 and 3, which are comprised of students whose parents decided for a lower track, those with a teacher recommendation for a low track scored 14 IQ-points lower than those with a high track recommendation. When comparing the students' GPA, the pattern was similar. Those students with a high track recommendation were achieving significantly higher, whether they were to attend the high track or not (Group 1: $M[SD] = 5.34[.37]$, Group 3: $M[SD] = 5.10[.42]$), than those without a recom-

mendation for the high track (Group 2: $M[SD] = 4.19[.31]$, Group 4: $M[SD] = 4.71[.47]$; mean separation tests with Tukey's HSD).

The students' age upon school entrance did not differ between the groups, $F(3, 511) = 1.42, p = .237$. However, at the time of recommendations and enrollment decisions, differences in age were found, $F(3, 521) = 9.92, p < .001$. Post-hoc analysis revealed that the effect referred to mean differences between those who followed the teacher recommendations. Students with a recommendation for and enrollment in the high track were significantly younger than students with a recommendation for and enrollment in the lower tracks (Group 1: $M[SD] = 119.74[4.88]$, Group 2: $M[SD] = 122.62[6.32]$). The same pattern was found for cultural capital, $F(3, 583) = 6.29, p < .001$, revealing those with a recommendation for and enrollment in the high track possessed significantly more books than students with a recommendation for and enrollment in the lower tracks (Group 1: $M[SD] = 66.93[54.75]$, Group 2: $M[SD] = 43.37[70.61]$).

There was a significant difference between the groups regarding the level of parental education, $H(3) = 90.348, p < .001$, with a mean rank of 328 for Group 1, 313 for Group 4, 264 for Group 3, and 208 for Group 2. 47.2% of those parents, who "down"-tracked their child, had attended a lower track themselves, whereas 59.5% of those parents, who "up"-tracked their child, had attended the high track.

6.6. Discussion

By recommending a secondary school, teachers influence their students' academic career. This study examined the impact of students' individual background variables (gender, age, migration background, learning disorders), students' family background variables (parents' educational level, cultural capital), and indicators for previous educational history and academic potential (GPA, grade retention or acceleration, IQ) on teachers' secondary school track recommendations. By doing so, we replicated recent findings regarding the teacher recommendations. In addition, we extended those findings by examining impact of learning disorders, and the correspondence of teacher recommendations and parents' decisions for final enrollment.

6.6.1. Explaining teacher recommendations

Overall, we were able to explain a lion's share of variance in teacher recommendations when predicting teacher recommendations in a comprehensive model considering both, primary and secondary effects. Only the primary effects were directly related to teacher recommendations. In line with previous findings, GPA showed the strongest relation with teacher recommendations, followed by TJs of cognitive ability and academic motivation, and students' intelligence (for GPA: Baumert et al., 2010; for TJs: Anders et al., 2010; for intelligence: Roos et al., 2011). Anders et al. (2010) found TJs of social behavior to be negatively related to track recommendation, i. e., the higher the judgement, the lower the recommendation, however, our results indicated a negligible effect in the comprehensive model.

As for the secondary effects on the recommendations, most background variables were significantly related to the primary effects under investigation. Therefore, teacher recommendation was influenced indirectly. Students' GPA as well as TJs of all dimensions were in favor of those more intelligent and younger. As intelligence is one of the most important single predictors of academic achievement (Roth et al., 2015; Spniath et al., 2006) and TJs are closely related to the students' GPA (Anders et al., 2010), these findings are not surprising. Regarding the advantage of younger students in GPA and all TJs, it is most likely due students, who have repeated a grade voluntarily or by advice and are consequently older than those who haven't (Bellenberg, 2005), because early school entrance and grade acceleration were uncommon ($< 1\%$).

GPA, TJs of cognitive ability, and TJs of academic motivation as primary effects on the track recommendation, were additionally influenced by parental education and learning disorders, respectively. The strong influence of parents' educational background on TJs and GPA replicates previous findings (e. g., Anders et al., 2010; Maaz & Nagy, 2009). In addition to the parents' educational background, the number of children books in the household predicted GPA significantly, and TJs of cognitive ability and academic motivation marginally. Glock et al. (2013) discuss cultural capital controversially, because teachers do not necessarily know about the number of books in the students' household. However, cultural capital operationalized through number of children books in the household is related to the SES and explicated through status-related indicators perceptible to teachers. Therefore, its predictive value is plausible. If a student was diagnosed with dyslexia, dyscalculia, ADHD, or any combination of those disorders, he or she obtained lower grades and was judged less intelligent and motivated. Although students

with dyslexia and dyscalculia are entitled for compensations in exams to prevent disadvantages in grading, reading and writing skills have been found to be crucial for school achievement (Gut, Reimann & Grob, 2015). Therefore, a lower GPA might be explainable, especially because students with ADHD are not granted any compensation. However, differences in TJs of cognitive ability, however, suggest that their judgment is biased, as it was controlled for actual ability. Teachers are mostly unprepared for teaching students with learning disorders, as courses on inclusive education are only obligatory in less than half of the university teacher education (Rischke, Baedorf & Müller, 2015). Consequently, teachers did not seem to discriminate between intellectual potential and the less evolved partial performance, as teachers seem to be unaware of them being (mostly) distinct.

Furthermore, the students' gender was positively related with GPA, replicating findings that girls obtain higher grades (e. g., Voyer & Voyer, 2014). Additionally, girls were judged higher on academic motivation and social behavior, also replicating previous findings (e. g., Anders et al., 2010; Legewie & DiPrete, 2012). Because no standardized tests of school achievement, nor assessment tools for the students' motivation and social behavior were administered, we may not conclude whether these differences in judgment are valid or not. Regarding mother tongue, differences in TJ of cognitive ability point toward a bias, because actual cognitive ability was controlled for and students whose mother tongue were not German were judged significantly less intelligent. However, the impact was very small and previous studies reported equal opportunities of being recommended for the high track, if students with migration background performed on the same level as those without (e. g., Glock et al., 2013; Kristen, 2006).

Overall, the secondary school track decision itself seemed less impacted by social background variables than by the teacher judgments and the students' GPA (Meulemann & Relikowski, 2016). Thus, the students' individual and parental background influences the students' academic career indirectly but profoundly (Dumont et al., 2014).

6.6.2. The relationship of teacher recommendation and parental enrollment decision

In line with previous findings, the majority of students enrolled in the track recommended by the teacher (Maaz & Nagy, 2009; Jonkmann et al., 2010). Only 7% of the students enrolled higher and 6% enrolled lower than the teacher recommendations, replicating the findings of Ditton and Krüsken (2009) for the federal states of Saxony and Bavaria almost exactly.

Secondary schools aim at rather specific educational and professional paths (Maaz et al., 2010) and bring about different achievement gains (e. g., Becker, Lüdtke, Trautwein & Baumert, 2006). "Down"-tracked students outperformed their prospective classmates by almost one standard deviation of the IQ-distribution in the administered intelligence tests. Therefore, those students might be disadvantaged in lower track secondary schools, as the learning environment might not match their intellectual needs. For those who attend a track higher than recommended, Schuchart and Weishaupt (2004) found, one third to graduate without grade retention while the remaining either switched to a lower track, dropped out, and a minority graduated after repeating a year. Others point toward the direction of "up"-tracked students benefiting from a challenging learning environment (e. g., Scharenberg, Gröhlich, Guill & Bos, 2010).

Overall, if recommendation and enrollment decision diverge, the teacher recommendation seems more closely related to the students' actual and potential performance. When comparing the "up"– and "down"–trackers' parental education, results point in the direction of reproducing the parents' own habitat, by enrolling the students in those tracks the majority of their parents had attended themselves. To clarify the impact of the social background on and the validity of mere parental enrollment decisions, it may be worthwhile to focus on those students, who attend a track not matching the original recommendation, and their achievement gains across schooling.

6.7. Limitations and outlook

The results reported in this article are based on a cross-sectional design. Thus, findings cannot be interpreted causally. Moreover, our sample is restricted to two federal states in Germany. Although we assessed the parental education and cultural capital as proxies for the socio-economic status, we did not assess the latter itself due to supervisory boards' regulations.

For further research, it may be worthwhile to examine children over a longer period of time. The educational career does not start in Grade 1 of primary school and the transition from primary to secondary school is a result of a longitudinal development (Ditton & Krüsken, 2009). While the choice of a kindergarten is dependent of the availability of a place and the parents' choice, the choice of a primary school is usually dependent of the parents' place of residence. Socio-economic status has an impact of the place of residence, since the higher the income, the higher the selection options and therefore, social disparities do

not only arise within the students' primary school class but between the schools as well (e. g., Maaz et al., 2010).

6.8. Conclusion

The present study sheds light on the continuous importance of social background on students' educational career and extends findings on the importance of TJs in decision-making procedure. The results of comparing those who obtain a high track recommendation with those who do not, do not necessarily question recommendation practices itself, for the practices seem mainly achievement driven as has been shown and concluded before (e. g., Ditton & Krüsken, 2009). Nevertheless, TJs and GPA have been shown to be impacted by the students' social background, the important role of the latter in the transition from primary to secondary school might need to be re-evaluated. In line with Meulemann and Relikowski (2016) most recent study on changes within the relative importance of primary and secondary effects, we conclude, that it seems necessary to (continuously) educate teachers on fair and accurate grading and formation of students' impressions, as TJs and GPA are closely related. Although parents mostly seem to follow teacher recommendations, in those cases they do not, it remains unsettled whether it is for the students' benefit or their reproduction of social habitat. Especially in those cases, where parents target a secondary school lower than the recommendation and students are diagnosed with learning disorders, counseling seems to be in order to allow each student to pursue the educational path he or she is most likely to perform at his or her best.

Table 6.2.: Description of student sample

		Total Sample	Recommendation for high track			Students whose parents *DID* follow teacher recommendation		Students whose parents *DID NOT* follow teacher recommendation	
			Yes	*No*	*Cohen's d*	*Group 1: High track*	*Group 2: Lower track*	*Group 4: "Down"-Tracker*	*Group 5: "Up"-Tracker*
		597	300	297		264	255	36	42
Gender (% female)		49.9	56.3	43.4		57.2	43.1	50.0	45.2
Age in months ($M(SD)$)		121.10 (5.87)	119.80 (4.88)	122.37 (6.46)	.38	119.74 (4.88)	122.62 (6.32)	120.15 (4.98)	120.81 (7.17)
School entrance age in months ($M(SD)$)		76.84 (4.39)	76.48 (4.18)	77.22 (4.58)	.16	76.40 (4.25)	77.23 (4.35)	77.00 (3.67)	77.13 (5.91)
Mother tongue (%)	1	79.1	84.3	73.7		83.7	76.5	88.9	57.1
	2	7.9	7.0	8.8		7.2	8.6	5.6	9.5
	3	13.1	8.7	17.5		9.1	14.9	5.6	33.3
Learning disorder (%)		10.6	2.3	18.9		1.9	20.8	5.6	7.1
GPA		4.80 (.73)	5.31 (.38)	4.27 (.62)	1.76	5.34 (.37)	4.19 (.31)	5.10 (.42)	4.71 (.47)
IQ (THINK)		103.10 (14.75)	111.45 (11.95)	94.68 (12.34)	1.33	111.96 (11.83)	94.01 (12.61)	107.71 (12.30)	98.75 (9.70)
Grade retention (%)		7.0	1.0	13.1		1.1	14.1	–	7.1
Grade acceleration (%)		.7	1.3	–		1.5	1.6	–	–
Parental education (%)	1	11.2	4.7	17.8		3.4	20.0	13.9	4.8
	2	30.8	22.3	39.4		20.8	42.0	33.3	23.8
	3	48.7	64.0	33.3		65.9	29.0	50.0	59.5
Cultural capital child (# of children books, $M(SD)$ *and range*)		54.90 (62.88) 0 – 1000	64.61 (53.78) 3 – 500	45.10 (69.63) 0 – 1000	.32	66.93 (54.75) 3 – 500	43.37 (70.61) 0 – 1000	47.92 (43.28) 10 – 200	55.73 (63.07) 10 – 400

Note. GPA refers to the mean grade in German, Math and General Studies. Gender: 1 = *female*, 2 = *male*. Mother tongue: 1 = *German*, 2 = *bilingual, i.e., German and one other language*, 3 = *other*. Parental education: 0 = *lowest track*, 1 = *medium track*, 2 = *highest track*. Teacher recommendation: 0 = *lower track*, 1 = *high track*.

Table 6.3.: Intercorrelations of measures

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	GPA										
(2)	IQ (THINK)	.618***									
(3)	Gender	−.083*	.008								
(4)	Mother tongue	.117**	.130**	−.003							
(5)	Age	−.282***	−.208***	.070†	−.052						
(6)	Parental education	.361***	.289***	−.034	.010	−.221***					
(7)	TJ_cognitive ability	.799***	.649***	.016	.141***	−.275***	.368***				
(8)	TJ_academic motivation	.686***	.487***	−.207***	.056	−.248***	.290***	.723***			
(9)	TJ_social behavior	.324***	.233***	−.188***	−.013	−.219***	.166**	.375***	.604***		
(10)	Teacher recommendation	.719***	.569***	−.129**	.130***	−.227***	.348***	.715***	.628***	.282***	
(11)	Parental enrollment decision	.655***	.490***	−.116**	.025	−.211***	.396***	.582***	.519***	.285***	.739***
Partial correlations											
(10)	Teacher recommendation (control GPA)		.280***	−.098*	.115*	−.089*	.166***	.270***	.245***	.080†	
(10)	Teacher recommendation (control IQ)	.464***		−.147**	.124**	−.120**	.220***	.462***	.428***	.124**	
(10)	Teacher recommendation (control GPA and IQ)			−.116**	.114*	−.071	.143**	.214***	.225***	.064	
(11)	Parental enrollment decision (control teacher recommendation)	.210***	.133**	−.52	−.115*	−.064	.222***	.115**	.156**	.070	
(11)	Parental enrollment decision (control GPA)		.204***	.099*	−.029	−.086†	.253***	.161***	.186***	−.075†	.569***

Note. GPA refers to the mean grade in German, Math and General Studies. Gender: 1 = *female*, 2 = *male*. Mother tongue: 1 = *German*, 2 = *other*. Parental education: 0 = *lowest track*, 1 = *medium track*, 2 = *highest track*. Teacher recommendation: 0 = *low tracker*, 1 = *high track*.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Table 6.4.: Results from logistics regressions predicting teacher secondary track recommendation

	Primary effects			**Secondary effects**		
Criteria:	***Teacher recommendation***			***Teacher recommendation***		
	β	*SE*	*Odd ratio*	β	*SE*	*Odd ratio*
Student background						
Gender				−.114*	.048	.60
Age				−.167***	.045	.94
Mother tongue				.128**	.037	2.01
Learning disorders				−.324***	.054	.10
Family background						
Parental education				.305***	.055	2.62
Cultural capital				.074*	.037	1.18
Teacher judgments						
Cognitive ability	.314***	.070	5.30			
Academic motivation	.229**	.072	3.20			
Social behavior	−.071	.047	.71			
GPA	.422***	.070	8.08			
IQ (THINK)	.118**	.044	1.84			
Nagelkerkes Pseudo-R^2:	**.683*** **			**.340*** **		

Note. GPA refers to the mean grade in German, Math and General Studies. Gender: 1 = *female*, 2 = *male*. Mother tongue: 1 = *German*, 2 = *other*. Parental education: 0 = *lowest track*, 1 = *medium track*, 2 = *highest track*. Teacher recommendation: 0 = *lower track*, 1 = *high track*. As we ran the logistic regression with M*plus*, the R^2 for binary outcomes refers to the latent response variable and not the observed binary outcomes, therefore it lacks expressiveness. To obtain information on explained variance, we ran analyses with SPSS 21 with class-wise standardized variables – betas, standard errors and significance were comparable throughout all analyses.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Table 6.5.: Results of logistic path model predicting teachers' secondary track recommendation through primary and secondary effects

	Primary and secondary effects										
Criteria:	***GPA***		***TJ_cognitive ability***		***TJ_academic motivation***		***TJ_social behavior***		***Teacher recommendation***		
	β	*SE*	β	*SE*	β	*SE*	β	*SE*	β	*SE*	*Odd ratio*
Student background											
Gender	−.056*	.028	.036	.030	−.186***	.037	−.177***	.044	−.060	.038	.54
Age	−.128***	.032	−.119***	.031	−.119***	.033	−.154***	.044	−.000	.042	1.00
Mother tongue	.043	.032	.062*	.031	−.004	.030	−.042	.033	.047	.031	1.83
Learning disorders	−.211***	.034	−.137***	.037	−.165**	.050	−.066	.046	−.035	.037	.55
Family background											
Parental education	.152***	.035	−.164***	.032	−.119**	.038	.071	.054	.060	.039	1.56
Cultural capital	.076**	.026	−.042†	.023	−.052†	.030	−.019	.032	−.009	.019	.95
Teacher judgments											
Cognitive ability									.311***	.069	5.40
Academic motivation									.222**	.065	3.14
Social behavior									−.084†	.045	.66
GPA									.398***	.063	7.48
IQ (THINK)	.485***	.027	.532***	.029	.385***	.035	.175***	.047	.105*	.051	1.73
R^2:	**.488*** **		**.492*** **		**.344*** **		**.127*** **				
Nagelkerkes Pseudo-R^2:										**.706*** **	

Note. Student and family background variables as well as TJs with GPA were intercorrelated (see Table 6.3, page 79). GPA refers to the mean grade in German, Math and General Studies. Gender: 1 = *female*, 2 = *male*. Mother tongue: 1 = *German*, 2 = *other*. Parental education: 0 = *lowest track*, 1 = *medium track*, 2 = *highest track*. Teacher recommendation: 0 = *lower track*, 1 = *high track*. As we ran the logistic regression with M*plus*, the R^2 for binary outcomes refers to the latent response variable and not the observed binary outcomes, therefore it lacks expressiveness. To obtain information on explained variance, we ran analyses with SPSS 21 with class-wise standardized variables – betas, standard errors and significance were comparable throughout all analyses.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

7. General discussion

The results of the conducted studies have been discussed in detail in the respecting chapters (see Sections 4.6, 5.6, and 6.6). Therefore, this discussion will only summarize the major findings and discus them in reference to accuracy of judgment, sources of bias, and consequences of (mis-)judgment. After that, the strengths and limitations of the present dissertation will be discussed and implications for further research as well teacher education will be drawn.

7.1. Diagnostic competencies of teachers

The dissertation focused on teacher judgments as part of their diagnostic competencies. The judgment of student characteristics, performance and behavior is a major part of a teachers' professional conduct (Schafer, 1993). While Study 1 examined teacher judgment accuracy, Studies 2 and 3 investigated the consequences of teacher judgment, strictly speaking, teacher referrals for gifted programming and teacher track recommendations for secondary schooling. A variety of sources of bias were included in all three studies. Although the main focus of the paper, which is either on accuracy or on consequences, is clearly identifiable, all three studies were tangent to the other focus as well.

7.1.1. Accuracy of judgment

The main results of Study 1 were, that accuracy levels differed in reference to the applied statistical method and that teachers were found to misjudge their students by 12 IQ-points on average. The systematic pattern of inaccuracy revealed, teachers to particularly misjudge students at either extreme of the IQ distribution. The assessed teachers tended to judge within the average level and underestimated those students, who are more intelligent, while overestimating those scoring lower on the intelligence test. This pattern is referred to as the central tendency error. Machts et al. (2016) discussed the possibility

of teachers being biased by the students' achievement in addition to cognitive biases. If a student was found to be high achieving, he or she was more likely to be judged beyond his or her actual cognitive ability. Results of Study 3 pointed toward the same directions, as the relation between GPA and TJs of cognitive ability was higher than the one between IQ and TJs of cognitive ability. Hanses and Rost (1998) reported that cognitive ability was not necessarily congruent with school achievement. Those students who achieve less than what could be expected based on their cognitive ability, i. e., underachiever, would be therefore systematically underrated. In fact, Baudson, Wollschläger, and Preckel (2014) compared gifted underachievers to average intelligence underachievers and gifted achievers, who are students that achieve at least as well as one would expect based on the results of an intelligence test. While teachers were found to differentiate between gifted underachievers and average intelligent underachievers correctly regarding their judgment of their students' cognitive ability, they seemed to have been misled by the students' achievement when comparing gifted underachievers and gifted achievers. Gifted underachiever were judged to be significantly less intelligent, although they performed at the same level as gifted achiever in the intelligence test. In other words, less achieving students are less likely to be considered (highly) intelligent. Those who cannot put their potential to performance need particular guidance (Reis & McCoach, 1986). If, however, the potential is not detected, the needed help may not be given. This is exemplified in higher rates of grade retention, school dropout and risk of unemployment for underachiever (e. g., Hillenbrand & Ricking, 2011).

Stang and Urhahne (2016) found in their study on TJA of achievement, attention, work habits and social behavior, that TJs of one characteristic were more explained through TJs of other characteristics than through information on the student. The authors discussed this finding as evidence of how teachers' personal opinions on and perceptions of student behavior form a generalized student picture. Anders et al. (2010) found teachers to differentiate between cognitive, motivational, and social behavior. Results showed teachers not to differentiate between cognitive ability and achievement as both scales loaded on the same factor "talent and achievement". In Study 2, confirmatory factor analyses showed teachers to judge students on related but distinct facets, namely cognitive ability, motivation, work attitude, and creativity. The results of Study 3 revealed TJs of cognitive ability and motivation to be closely related, while their judgment of social behavior seemed rather distinct. The present dissertation may not answer the questions, whether teachers form a distinct or general perception of a students, however as TJs are interrelated and previous studied reported similar findings, it may be assumed that TJs

of all kinds of student characteristics are highly related and as discussed above mainly performance-driven.

As for the accuracy rates themselves, Study 1 replicated findings of the recent meta-analytic findings of by Machts et al. (2016). The relation between TJs and IQ revealed large effects and approximately one-third of the shared variance was found. Machts et al. (2016) considered the moderate mean judgment accuracy as acceptable, with a reference to the fact, that intelligence in comparison to achievement as a rather distal student characteristic that "requires stronger inferences than the judgment of achievement" (p. 99). However, one-third of shared variance means that two-thirds were not explained, leaving much room for the incremental influence of other variables. The large effect of the relation between TJs and IQ skews the picture of accuracy, as in fact, teacher judgments were of by approximately one standard-deviation of the IQ-distribution. These findings educe the question of quality criteria of teacher judgment accuracy. The accurate judgment of cognitive ability becomes necessary to teachers, when identifying students' special needs in either extreme of the intelligence distribution and predicting which students are academically suitable for higher education and those who are not. As findings point towards the direction of improvable judgment accuracy (DeYoung, 2008; Machts et al., 2016, see also findings of Study 1), any decisions impacting a student's educational trajectory should be accompanied by standardized tests of achievement and potential. Nevertheless, teachers have an important part in identifying those students with special needs as they gather information over a long time and in various educational settings (e. g., Baudson, 2010). Therefore, while practitioners should remember teachers are a valuable source of information and extend the teachers' perception by standardized tests satisfying psychometric quality standards, researchers should interpret large effects more carefully as they still leave room for biasing influences.

7.1.2. Sources of bias

Study 2 and 3 revealed, the TJ of cognitive ability to be important for their nomination of students for gifted programming and for the secondary track recommendation at the end of Grade 4. Furthermore, their judgment of creativity was important in nominating students for the enrichment program, while their judgment of academic motivation and work conduct was found to be important for secondary school track recommendations. Thus far, the results seemed plausible and desirable: The more intelligent a teacher perceive a student to be, the more likely this student was recommended for enrichment program or

the highest secondary track, likewise for creativity (Study 2) and motivation (Study 3). However, TJs were not independent of student characteristics. In Study 2, the parental education affected all TJs, for girls even the nomination probability directly. In Study 3, the students' background was considered in more detail. TJs were affected by the students' gender, his or her age, whether he or she had learning disorders, and the parental education. Grades were even more impacted than TJs of achievement-related student characteristics. In Study 1, the distance of judged and actual IQ was partly explained by the students' mother tongue and the parents' educational background, i. e., if the actual IQ was misjudged, the inaccuracy could be explained by students' mother tongue and the parents' educational background to some extent.

The parental education was considered in all three studies and has been found to be important. Although parental education is only a proxy for SES, results substantiate its important impact on students' educational career. In 2000, PISA results eventuated in shock waves in Germany for two reasons. For one, the students themselves performed within the OECD average level. For another, the relation between students' social background and educational success was higher than in every other participating country (OECD, 2001). Ehmke and Jude (2010) compared the relation of social background and achievement over four PISA assessments. Results indicated that the social gradient decreased and settled within the OECD average range. Nevertheless, the authors concluded that the linkage of SES and education still needs to be uncoupled. The results of the present studies indicated that students from an educated home are judged more intelligent than they actually are (Study 1) and especially for girls, the family background was found to be important when nominated for gifted education. However, teacher judgments themselves seemed to be more affected by the family background than the nomination probability and secondary track recommendations. The strong impact has been studied and found in many studies in Germany (e. g., Baudson et al., 2014; Maaz & Nagy, 2009; OECD, 2011; Prenzel et al., 2013). As a comparison, Finnish students have been found to perform significantly above OECD average and the linkage between SES and performance was significantly below average (Välijärvi & Malin, 2003). In Finland, students attend a comprehensive nine-year compulsory school, which is considered to be one reason for almost negligible social divisions and structural inequality in primary and secondary education (Sahlberg, 2007). In Germany, students are tracked at the age of ten, after four years of learning. However, the high linkage of student achievement and parents' SES found at the age of fifteen, when PISA takes place, have been reported in international achievement tests in primary school students already (Bos et al., 2007; Arnold et al., 2007). Therefore, tracking alone might

intensify but not cause social disparities solely (Ditton, Krüsken & Schauenberg, 2005; Maaz et al., 2009; Trautwein, 2014).

Closely related to the parents' SES is the migration background. In Germany, students with migration background typically grow up in low(er) SES households (Diefenbach, 2005). Furthermore, students with a migration background perform lower in national and international achievement tests than students without migration background (Stanat, Rauch & Segeritz, 2010). Even when controlling for the families' SES, students performed lower in achievement tests (Stanat & Christensen, 2006). In the present studies, the migration background was operationalized through the students' mother tongue. Previous studies have found that language skills were more essential than the migration background itself, as students with migration background were not disadvantaged if they performed at the same level as students without migration background (Kristen, 2006; Schründer-Lenzen & Merkens, 2006). In Study 1, results indicated that students whose mother tongue was not German were judged to be significantly less cognitively able. In Study 3, findings point in to the same direction. However, only TJs of cognitive ability were predicted through mother tongue, TJs of motivation and social behavior, as well as the students' GPA seemed independent. For GPA, these findings undermine previous findings that, if students perform at the same level, migration background is insignificant (e. g., Kristen, 2006; Glock et al., 2013). Regarding motivation, Stanat, Segeritz, and Christensen (2010) found students with migration background to be as motivated to educationally succeed as those without migration background, therefore, equal TJs of motivation seemed accurate. Data analysis in Study 3 controlled for actual ability, the negative impact of mother tongue in TJs of cognitive ability hint at a potential bias. Even when being equally intelligent, those students whose mother tongue was not German where judged less able. This judgment error could lead to a systematic disadvantage of students from a cultural- and language-divers background. Furthermore, as migration background and socioeconomic status are related, students could be double unprivileged as discussed in Study 1 (see Subsection 4.6.3).

Findings of Study 3 revealed students with a learning disorder to be judged less intelligent after controlling for actual differences in intelligence, as well as less motivated and they received lower grades. While results indicated a systematic underestimation of the students' cognitive ability, it remains unsettled whether students actually showed less motivated behavior. Regarding the lower GPA, students with dyslexia and dyscalculia have a right to contextual and procedural compensation in exams, which is granted by the German basic law. Hence, no students should be at a disadvantage because of dyscalculia

and/or dyslexia, but should be given the same chance of educational development, when achievement level is comparable. However, Marwege (2013) summarized that students would rather not take advantage of those compensation for fear of being stigmatized and those who ask for rightful compensation do not receive in the amount they should, because of obstacles in school policies. Emmer and Stough (2001) reviewed findings on classroom management in the light of the inclusion of children with special needs. They concluded teachers rather to be unprepared for the task because of lacking knowledge and skills on learning disorders. For the United States, Kantor (2011) found 98% of teachers felt prepared for teaching regular education students, while depending on the learning disorder or special need up to 43% (strongly) disagreed to being prepared. According to Rischke et al. (2015), less than half of the universities in Germany included obligatory courses on inclusive education. It may be assumed, that German teachers would feel unprepared, too. In addition to knowledge, the positive attitude toward teaching students with learning disabilities is a prerequisite for inclusive education (Loreman, Earle, Sharma & Forlin, 2007). Jung (2007) found teachers to be more confident and report a more positive attitude to teaching children with learning disabilities, when being substantially educated on the matter. In other words, as long as teachers feel unprepared for the task because of lacking education, they are less likely to acquire a positive attitude, which has been found to be necessary for successful inclusive education (Loreman et al., 2007). Although the systematic disadvantage of students with learning disorders is an indication of biases in teacher judgments, other studies hint at a severe lack of education on the topic in teacher education programs. Therefore, teachers may not be aware of the distinction of intellectual potential and, e. g., actual writing skills.

Regarding gender, findings have been inconclusive and whether teachers systematically favor one or the other could not be answered with the present dissertation either. However, the picture might become clearer: Study 1 did not reveal a systematic influence on TJA, neither findings pointing toward a favor of boys (Trautwein & Baeriswyl, 2007) nor girls Mullola et al. (2012) were replicated. On the contrary, findings point towards an unbiased judgment of cognitive ability regarding gender, as in fact, group differences were not to be expected (Brunner et al., 2008). Study 2 indicated a different nomination process for girls that included the parental education to their favor and participation in class to their disadvantage. Hypothetical explanations were discussed in Subsection 5.6.2. Study 3 replicated findings on girls receiving better grades (Voyer & Voyer, 2014) and found TJs of academic motivation and social behavior to favor girls in line with previous findings (Anders et al., 2010; Legewie & DiPrete, 2012). As neither study included standardized tests of achievement, motivation nor social behavior, it may not be settled

whether these differences in judgment are because of actual group differences or because of teachers' personal opinions. Wentzel (1993) found teachers' preferred social and academic behavior to be closely related to the students' GPA, as well as social and academic behavior to be closely related themselves. As discussed above, TJs are related to the students' achievement. Whether girls performed at a higher level and were, hence, judged to be more motivated and socially behaving or actually do show more social and academic behavior may not be settled.

7.1.3. Consequences of (mis-)judgment

Teacher judgments were found to lack accuracy and to be biased by the students' social background. Nevertheless, teachers were found to consider educational justice to be highly important and endeavor fair and just judgments (Dalbert, Schneidewind & Saalbach, 2007). Furthermore, Popham (2009) concluded, that assessment literary was necessary for the "long-term well-being [of teachers and] the educational well-being of their students" (p.11). Lacking knowledge in classroom assessment skills, may be considered disadvantageous for the teachers and the students. Although teachers were given a questionnaire assessing some background variables, it did not exceed their age, gender, familiarity with the class and years of professional experience. The impact of classroom assessment literacy and their respective skills may not be investigated. The consequences of misjudgment are therefore discussed only in the light of their students' educational development.

Study 2 investigated the nomination probability for students to participate in gifted education. Although biases in TJs have already been discussed, the different nomination approach with respect to girls and boys is noteworthy. Girls and boys were equally likely to be nominated by their teachers, replicating meta-analytic findings by Petersen (2013), but higher TJs of participation in class led to lower probability of nomination for girls. An explanation could be the attribution of success to innate ability in boys and to hard work in girls (Fennema et al., 1990; Siegle & Reis, 1998). Petersen (2013) found boys to be more likely to participate in gifted education programs. Although the nomination rates were the same, the differences in underlying components hint at a possible disadvantage for those girls, who participate voluntarily and regularly. Instead of interpreting frequent participation as an indicator of giftedness, it seems understood as the compensation for lacking ability. As most programs aim at an adequate balance of gender ratio and teach-

ers often function as a gatekeeper, the teachers' approaches in selecting suitable students should be further understood.

Study 3 revealed most students to follow the teachers' recommendation of secondary school track. As they differentiate between those assumable successful in higher education and those who are most likely not to succeed in the highest track, their recommendation is of consequence (Bellenberg, 2012). While the secondary track recommendation was less affected by the students' background, the judgments of achievement and student characteristics were profoundly influenced. Potentially successful students might therefore not be recommended for the highest track. Although students can switch between tracks or change to a higher track after graduating a lower track, they seem disadvantaged as tracks represent different learning environments and bring about different achievement gains (Becker et al., 2006; Baumert, Stanat & Watermann, 2006).

Grades are the societies' operationalization of the educations' selection purpose, as high achieving students are preferably chosen by universities and employers (e. g., Trapmann, Hell, Weigand & Schuler, 2007). Furthermore, low achieving students face the danger of grade retention or switching to a lower secondary track, if certain grade criteria is not met (e. g., Jimerson, 2001). Therefore, the comparability of grades between students of different classes, years and schools is one of their fundamental necessities (Sacher, 2009). However, teachers tend to use their specific classroom as a frame of reference for assigning grades (e. g., Trautwein & Baeriswyl, 2007). As the class ability level is used as the comparison standard a student is related to, it should be accurately judged. Although on average teachers were found to judge it rather accurately, inter-individual differences in TJA were found (see Subsection 4.5.2). If the overall class's ability level is misjudged in either direction, the classroom as a frame of reference becomes questionable with regard to comparability of judged achievement or assigned grades over classes (Tarelli, Valtin, Bos, Bremerich-Vos & Schwippert, 2012). An example of this is when a student is judged too strictly by overestimating or too liberally by underestimating the classes' ability forcing reference-group effects to increase.

As a consequence, the student judged too strictly may not be admitted to his or her university of choice, if he or she gained lower grades in comparison to students from other classes. Trapmann et al. (2007) conducted a meta-analysis on the validity of school grades for academic achievement at the university. Although grades are often criticized to be unreliable, invalid, and subjective, the authors found students' GPA in school to significantly predict students' GPA at undergraduate and graduate university level. The mean corrected validity coefficients ranged from .26 to .56, which indicate large effects. In the

search for more appropriate selection criteria, a different meta-analysis, conducted by Hell, Trapmann, Weigand & Schuler (2007a), examined the validity of admission interviews. Results indicated that structured interviews are more valid than unstructured interviews for academic success, but the incremental validity of interviews to school grades was found to be low. Subject-specific admission tests, however, were found to be as valid as school grades (Hell, Trapmann, Weigand & Schuler, 2007b). To conclude, whether teachers use a restricted frame of reference or not, school grades have a substantial validity for academic success, but standardized subject-specific admission tests could be a fairer but still equally valid admission criteria. Not only teachers should rethink their grading practices, society could reevaluate the grades' significance.

7.2. Strenghts, limitations and directions for further research

The specific strengths, limitations and implications of the three research articles have been discussed in the respective chapters (see Sections 4.7, 5.7, and 6.7), therefore, only those addressing overall aspects will be made subject of discussion. The present dissertation addresses an important topic in educational research and practice in general, and for elementary school in particular. The teachers' classroom assessment skills were investigated empirically from a broad interdisciplinary perspective integrating findings and theoretical models from educational psychology, pedagogic, and sociology. By being based on two large independent samples and analyzed by applying up to date methodological and psychometrical standards, the results form a comprehensive picture of teachers' diagnostic competencies regarding classroom assessment.

Although the piloting and standardizations studies of the Test for the Assessment of Intelligence in Childhood (Baudson & Preckel, 2013a; Baudson et al., 2016) assessed the students' abilities and background in a comprehensive way, information on teachers may be considered extensible. Future studies could include measures of teachers' general pedagogical and psychological knowledge (e. g., Voss et al., 2011) to draw conclusions in how far the professional knowledge is related to his or her diagnostic abilities. Also, parents' educational level merely represents a proxy variable to socio-economic status. A broader assessment of the students' family background could further shed light on the psychological processes behind its importance.

As both samples were German only and research questions were in parts restricted to particularities of the German school system, generalization of findings requires further investigation. Furthermore, both samples are based on a cross-sectional design. Longitudinal studies in elementary schools indicated that teachers form their opinions of and expectations for students' very early (e. g., Ditton & Krüsken, 2009). Additionally, Maaz and Nagy (2009) concluded that social disparities are the results of the interaction of students' social background and school grades, teacher recommendations, and parental enrollment decision. To untangle these interrelationships, longitudinal studies starting before and lasting longer than primary school seem necessary.

Rivkin, Hanushek, and Kain (2005) emphasized on the important role of teaching quality for school achievement, as they found students to exceed their age-mates up to one year in learning content, if taught by excellent teachers. Furthermore, the results of this education-economical study indicated that "moving one standard deviation up the teacher quality distribution" would be more efficient than costly school reform, e. g., reducing class size (p. 417). Despite the integration of psychological, pedagogical, and sociological models, methods, theories, and findings, an even more broadened perspective might help to further understand the necessities, possibilities, costs and feasibility of a schooling system from elementary to tertiary level of equal opportunities with the goal of fostering each student to fulfill his or her utmost potential.

7.3. Implications for teacher education

In comparison to German students, Finnish students have performed above average in many international standardized school achievement tests (e. g., Ehmke & Jude, 2010). Finnish teachers are very well educated and Finnish teacher education programs are outstanding in depth and complexity in international comparison (Westbury, Hansen, Kansanen & Björkvist, 2005). In Germany, up to two-thirds of university graduates of teacher education programs, who have worked for a maximum five years, stated to have felt unprepared to the demands of the teaching profession (Süßlin, 2012).

Sahlberg (2007) summarizes the following education policies as responsible for the success of the Finnish education system: *Flexibility and loose standards* instead of standardization, *broad learning combined with creativity* instead of focus on literacy and numeracy, and *intelligent accountability with trust-based professionalism* instead of consequential

accountability. Although differences are at hand and the school system matters, the individual teacher's ability has a great(er) impact (Lipowsky, 2006). One of the most popular and most disputed experiments regarding the impact of teachers took place in 2007 in Sweden. The documentary "Klass 9A" accompanied students in the transition from one of the worst performing classes to one of the top-three in the final state exam in five months. All teachers were replaced by those, whose students regularly performed best in state exams. By the end the "Klass 9A" was placed third in overall achievement nationwide and outperformed all other classes in Mathematics (Kucklick, 2011).

Jürges and Schneider (2008) approached the necessity of educated and well trained teachers from an economic perspective and found the benefits of educating teachers exceed the cost, as teachers might compensate disadvantages of children from less educated households. Teachers, therefore, should be better prepared for the tasks ahead (Süßlin, 2012). As reforms in teacher education tend to be protracted, a focus on what each and every teacher may do for themselves could be expedient.

The understanding of general pedagogical/psychological knowledge by Voss et al. (2011) includes the evaluation of the applied classroom assessment tools. An example of how assessment skills may be monitored and evaluated was given by Hesse and Latzko (2009). They proposed a 5-step procedure to self-assess diagnostic competencies based on Helmke (2007): (1) Choice of a student or test characteristic, for which the teacher would assess his or her judgment ability; (2) prediction of student behavior or performance on the test; (3) assessment of student behavior or test performance; (4) comparison of predicted and assessed behavior and test performance; (5) analysis of the results of the comparison and search for possible reasons for variance. If a teacher would like to assess his or her students' reading abilities, he or she would predict the students' performance and then test their reading abilities with a standardized test, compare the test results to the prior judgments and analyze any discrepancies for possible judgment errors. Furthermore, Hesse and Latzko (2009) recommend teachers to work in professional groups to compare teaching and assessment methods. Teacher education should enable teachers to assess the students' needs and abilities, as well as their own quality as a teacher in an informed manner, therefore they need to be knowledgeable and reflective in their professional knowledge (Wolf, 1993; Voss et al., 2011).

8. Conclusion

The aim of the present dissertation was to contribute to a further understanding of teachers' classroom assessment skills. All three studies revealed a rather problematic impact of achievement-unrelated student information on teacher judgments. As teacher judgments were found to be predictive of referral decisions, such as nomination for gifted programming and recommendation of a secondary track, the influence of achievement-unrelated student information on student's educational paths is alarming. Because teachers approximately spend half of their everyday work in assessment-related activities and their referral decisions influence every student's educational career, they should be prepared for the task of judgment and feel confident when assessing their students' achievement-related characteristics, academic potential, and performance. Although the necessities of changes in teacher education policies are clearly recognizable, teachers should be encouraged to enhance their knowledge and classroom assessment skills by evaluating their general pedagogical/psychological knowledge autonomously and/or in professional teams.

References

Achenbach, T. M. & Rescorla, L. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT, USA: University of Vermont.

Anders, Y., McElvany, M., & Baumert, J. (2010). Die Einschätzung lernrelevanter Schülermerkmale zum Zeitpunkt des Übergangs von der Grundschule auf die weiterführende Schule: Wie differenziert urteilen Lehrkräfte? [Teacher judgments of learning related student characateristics at the time of secondary school tracking]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 313–330). Bonn, Berlin, Germany: BMBF.

Arnold, K.-H., Bos, W., Richert, P., & Stubbe, T. C. (2007). Schullaufbahnpräferenzen am Ende der vierten Klassenstufe [Secondary school track aspirations at the end of forth grade]. In W. Bos, S. Hornburg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, & R. Valtin (Eds.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 271–297). Münster, Germany: Waxmann.

Artelt, C. & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften [Diagnostic competence of teachers]. *Zeitschrift für Pädagogische Psychologie*, *23*, 157–160. doi:10.1024/1010-0652.23.34.157.

Aukrust, V. G. (2008). Boys' and girls' conversational participation across four grade levels in Norwegian classrooms: Taking the floor or being given the floor? *Gender and Education*, *20*, 237–252. doi:10.1080/09540250802000413.

Baer, J. & Kaufman, J. C. (2008). Gender differences in creativity. *Journal of Creative Behavior*, *42*, 75–105. doi:10.1002/j.2162-6057.2008.tb01289.x.

Baker, J. A., Bridger, R., & Evans, K. (1988). Models of under-achievement among gifted preadolescents: The role of personal, family, and school factors. *Gifted Child Quarterly*, *42*, 5–14. doi:10.1177/001698629804200102.

Barber, C. & Tornay-Purta, J. (2008). The relation of high-achieving adolescents' social perceptions and motivation to teachers' nominations for advanced programs. *Journal of Advanced Academics*, *19*, 412–443. doi:10.4219/jaa-2008-813.

Barbu, S., Nardy, A., Chevrot, J.-P., Guellaï, B., Glas, L., Juhel, J., & Lemasson, A. (2015). Sex differences in language across early childhood: Family socioeconomic status does not impact boys and girls equally. *Frontiers in Psychology*, *6:1874*. doi:10.3389/fpsyg.2015.01874.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational and Behavioral Statistics*, *6*, 267–285. doi:10.2307/1164877.

Bates, C. & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, *21*, 177–187. doi:10.1080/01443410124450.

Baudson, T. G. (2010). Nominationen von Schülerinnen und Schülern für Begabtenfördermaßnahmen [Nominating students for gifted and talented programs]. In F. Preckel, W. Schneider, & H. Holling (Eds.), *Test und Trends – Diagnostik von Hochbegabung* (pp. 89–117). Göttingen, Germany: Hogrefe.

Baudson, T. G. (2011). *The (mis-)measure of children's cognitive abilities*. PhD thesis, University of Trier, Germany.

Baudson, T. G., Fischbach, A., & Preckel, F. (2014). Teacher judgments as measures of children's cognitive ability: A multilevel analysis. *Learning and Individual Differences*, -. doi:10.1016/j.lindif.2014.06.001.

Baudson, T. G. & Preckel, F. (2013a). Development and validation of the German test for (highly) intelligent kids – (T(H)INK). *European Journal of Psychological Assessment*, *26*, 171–181. doi:10.1037/t23598-000.

Baudson, T. G. & Preckel, F. (2013b). Intelligence and creativity: Their relation with special attention to reasoning ability and divergent thinking. Implications for giftedness research and education. In K. H. Kim, J. C. Kaufman, J. Baer, & B. Sriraman (Eds.), *Creatively gifted students are not like other gifted students: Research, theory, and practice* (pp. 205–236). Rotterdam, The Netherlands: Sense Publishers.

Baudson, T. G. & Preckel, F. (2013c). Teachers' implicit personality theories about the gifted: An experimental approach. *School Psychology Quarterly*, *28*, 37–46. doi:10.1037/spq0000011.

Baudson, T. G., Wollschläger, R., & Preckel, F. (2014). *Was unterscheidet überdurchschnittlich begabte von durchschnittlich begabten Underachievern?* [Differences between average talented and gifted underaciever](Arbeitsgruppe "Hochbegabung und Begabtenförderung", F. Preckel & M. Vock) Vortrag bei der 2. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF, 03.–05. März 2014), Frankfurt am Main, Germany.

Baudson, T. G., Wollschläger, R., & Preckel, F. (2016). *Test zur Erfassung der Intelligenz im Grundschulalter – THINK 1-4* [Test for assessment of intelligence in childhood]. Göttingen, Germany: Hogrefe.

Baudson, T. G., Wollschläger, R., Preckel, F., & Vock, M. (2014). Intellektuelle Hochbegabung [Intellectual giftedness]. In A. Lohaus & M. Glüer (Eds.), *Entwicklungsförderung im Kindesalter* (pp. 99–115). Göttingen, Germany: Hogrefe.

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competencies of teachers]. *Zeitschrift für Erziehungswissenschaft*, *9*, 469–520.

Baumert, J., Maaz, K., & Jonkmann, K. (2010). Der Übergang von der Grundschule in die weiterführende Schule als Forschungsgegenstand: Robuste Befunde, die Bewährung von Wert-Erwartungs-Modellen und offene Fragen [The transition from primary school to secondary school as object of research]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 313–330). Bonn/Berlin, Germany: BMBF.

Baumert, J., Maaz, K., & Trautwein, U. (2009). Bildungsentscheidungen [Educational decisions]. *Zeitschrift für Erziehungswissenschaft*, *Special Issue 12–09*, 153–182.

Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus [Differential achievement trajectories in a tracked school system]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen – differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 95–188). Wiesbaden, Germany: VS Verlag.

Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik: Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? [Achievement gains in mathematics: Evidence for differential achievement trajectories in

a tracked school system?]. *Zeitschrift für Pädagogische Psychologie*, *20*, 233–242. doi:10.1024/1010-0652.20.4.233.

Becker, R. (2016). Chancenungleichheit bei der Einschulung und in der Primarstufe. Theoretische Überlegungen und empirische Evidenzen [Equal opportunites at primary school entrance]. *Zeitschrift für Grundschulforschung*, *9*, 7–19.

Bellenberg, G. (2005). Wege durch die Schule - Zum Zusammenhang zwischen institutionalisierten Bildungswegen und individuellen Bildungsverläufen im deutschen Schulsystem [Educational paths in germany]. *Bildungsforschung*, *2*.

Bellenberg, G. (2012). *Schulformwechsel in Deutschland - Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I* [Transition between secondary school tracks in Germany]. Gütersloh, Germany: Bertelsmann Stiftung.

Berkemeyer, N., Bos, W., Manituis, V., Hermstein, B., & Khalatbari, J. (2013). *Chancenspiegel 2013 - Zur Chancengerechtigkeit und Leistungsfähigkeit der deutschen Schulsysteme mit einer Vertiefung zum schulischen Ganztag* [Opportunities in German school systems - 2013]. Gütersloh, Germany: Bertelsmann Stiftung.

Bianco, M., Harris, B., Garrison-Wade, D., & Leech, N. (2011). Gifted girls: Gender bias in gifted referrals. *Roeper Review*, *33*, 170–181. doi:10.1080/02783193.2011.580500.

Biesta, G. (2015). What is education for? On good education, teacher judgement, and educational professionalism. *European Journal of Education*, *50*, 75–87. doi: 10.1111/ejed.12109.

Biethahn, F. N. & Hogrebe, N. (2016). *"Wo haben Sie das denn gelernt?" – Eine Bestandaufnahme von Informationsquellen und Selbsteinschätzungen zur Notenvergabe bei Grundschullehrkräften* [Where from do elementary school teachers gather their information on grading?]. Vortrag bei den Abschlusspräsentationen der studentischen Projektarbeiten in der Forschungsorientieren Vertiefung des MSc-Tracks "Kompetenzentwicklung im Lebenslauf" (KEIL) unter der Leitung von F. Preckel, 11. Juli 2016, Trier, Germany.

Bildungsserver (2016). Entdeckertag in Rheinland-Pfalz [Gifted education in primary schools in Rhineland-Palatinate]. Retrieved from http://grundschule.bildung-rp.de/projekte/fruehe-hochbegabtenfoerderung/entdeckertag.html.

Birkel, P. & Pritz, V. (1980). Sprechflüssigkeit und Vorinformationen als validitätsmindernde Faktoren bei mündlichen Prüfungen [The influence of speech fluency and advance information on the validity of oral exams]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *12*, 284–289.

Borko, H., Roberts, S. A., & Shavelson, R. (2008). Teachers' decision making: from Alan J. Bishop to today. In P. Clarkson & N. Presmeg (Eds.), *Critical issues in mathematics education* (pp. 37–67). New York, NY, USA: Springer.

Bos, W., Schwippert, K., & Stubbe, T. C. (2007). Die Koppelung von sozialer Herkunft und Schülerleistung im internationalen Vergleich [The relation of school achievement and social background]. In W. Bos, S. Hornburg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, & R. Valtin (Eds.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 225–247). Münster, Germany: Waxmann.

Bos, W., Voss, A., Lankes, E.-M., Schwippert, K., Thiel, O., & Vaitlin, R. (2004). Schullaufbahnempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe [Teacher track recommendations for secondary schools at the end of fourth grade]. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin, & G. Walther (Eds.), *IGLU 2006 - Einige Länder der Bundesrepublik Deutschland im nationalen und internationalem Vergleich* (pp. 165–227). Münster, Germany: Waxmann.

Boudon, R. (1974). *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York, NY, USA: John Wiley & Sons.

Boudon, R. (2003). Beyond rational choice theory. *Annual Review of Sociology*, *29*, 1–21. doi:10.1146/annurev.soc.29.010202.100213.

Bourdieu, P. & Passeron, J.-C. (1971). *Die Illusion der Chancengleichheit* [The Illusion of equal opportunities]. Stuttgart, Germany: Ernst Klett Verlag.

Bourdieu, P. & Passeron, J.-C. (2007). *Die Erben: Studenten, Bildung und Kultur* [The heirs: Students, education and culture]. Konstanz, Germany: UVK.

Brimi, H. M. (2011). Reliability of grading high school work in english. *Practical Assessment, Research & Evaluation*, *16*. Available online: http://pareonline.net/getvn.asp?v=16&n=17.

Brown, S. W., Renzulli, J. S., Gubbins, E. J., Zhang, W., Siegle, D., & Chen, C. H. (2005). Assumptions underlying the identification of gifted and talented students. *Gifted Child Quarterly*, *49*, 68–79. doi:10.1177/001698620504900107.

Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, *36*, 403–421. doi:10.1016/j.intell.2013.05.009.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. doi:10.1037/0022-3514.42.1.116.

Carman, C. A. (2011). Adding personality to gifted identification: Relationships among traditional and personality-based constructs. *Journal of Advanced Academics*, *22*, 412–446. doi:10.1177/1932202X1102200303.

Carter, R. & Wojtkiewicz, R. A. (2000). Parental involvement with adolescents' education: Do daughters or sons get more help? *Adolescence*, *35*, 29–44.

Caspi, A., Chajut, E., & Saporta, K. (2008). Participation in class and in online discussions: Gender differences. *Computers & Education*, *20*, 718–724. doi:10.1016/j.compedu.2006.08.003.

Chen, F. F. (2007). Sensitivity of goodness-of-fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. doi:10.1080/10705510701301834.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/s15328007sem0902_5.

Clinkenbeard, P. R. (2012). Motivation and gifted students: Implications of theory and research. *Psychology in the Schools*, *49*, 622–630. doi:10.1002/pits.21628.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ, USA: Lawrence Earlbaum Associates.

Colman, A. M. (2009). *A Dictionary of Psychology* (3rd ed.). New York, NY, USA: Oxford University Press.

Cooper, H., Findley, M., & Good, T. (1982). Relations between student achievement and various indexes of teacher expectations. *Journal of Educational Psychology*, *74*, 577–579. doi:10.1037//0022-0663.74.4.577.

Copenhaver, R. W. & McIntyre, D. J. (2012). Teachers' perceptions of gifted students. *Roeper Review*, *14*, 151–153. doi:10.1080/02783199209553411.

Dai, D. Y., Swanson, J. A., & Cheng, H. (2011). State of research on giftedness and gifted education: A survey of empirical studies published during 1998–2010. *Gifted Child Quarterly*, *55*, 126–138. doi:10.1177/0016986210397831.

Dalbert, C., Schneidewind, U., & Saalbach, A. (2007). Justice judgments concerning grading in school. *Contemporary Educational Psychology*, *32*, 420–433.

Daniel, L. G. & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, *91*, 331–344. doi:10.1080/00220679809597563.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21. doi:10.1016/j.intell.2006.02.001.

Demaray, M. K. & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*, 8–24. doi:doi:10.1037/h0088969.

DeYoung, C. G. (2008). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge Handbook of Intelligence* (pp. 711–731). New York, NY, USA: Cambridge University Press.

Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G., & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning and Individual Differences*, *22*, 844–849. doi:10.1016/j.lindif.2012.04.004.

Diefenbach, H. (2005). Schulerfolg von ausländischen Kindern und Kindern mit Migrationshintergrund als Ergebnis individueller und institutioneller Faktoren [School success of foreign students and students with migration background as the results of individual and institutional influences]. In Arbeitsstelle interkulturelle Konflikte und gesellschaftliche Integration (Ed.), *Migrationshintergrund von Kindern und Jugendlichen: Wege zur Weiterentwicklung der amtlichen Statistik* (pp. 43–54). Bonn, Berlin, Germany: BMBF.

Diewald, M. & Riemann, R. (2014). Wie können wir die Ursachen sozialer Ungleichheit verstehen? [How do we understand the reasons for social disparities?]. In B. Spinath (Ed.), *Empirische Bildungsforschung. Aktuelle Themen der Bildungspraxis und Bildungsforschung* (pp. 68–78). Berlin Heidelberg, Germany: Springer.

Ditton, H. (2013). Wer geht auf die Hauptschule? Primäre und sekundäre Effekte der sozialen Herkunft beim Übergang nach der Grundschule [Who goes to the German "Hauptschule" (the lower ability band secondary school)? Primary and secondary effects of social origin in the transition to secondary school]. *Zeitschrift für Erziehungswissenschaften*, *16*, 731–749. doi:10.1007/s11618-013-0440-y.

Ditton, H. & Krüsken, J. (2009). Bildungslaufbahnen im differenzierten Schulsystem – Entwicklungsverläufe von Laufbahnempfehlungen und Bildungsaspirationen in der Grundschulzeit [Educational careers in a tracked school system—Development of teacher recommendations and educational aspirations over the elementary school years]. *Zeitschrift für Erziehungswissenschaften*, *Special Issue 12–09*, 74–102. doi:10.1007/978-3-531-92216-4_4.

Ditton, H., Krüsken, J., & Schauenberg, M. (2005). Bildungsungleichheit – der Beitrag von Familie und Schule [Educational disparities and the contribution of family and school]. *Zeitschrift für Erziehungswissenschaften*, *8*, 285–304. doi:10.1007/s11618-005-0138-x.

Döpfner, M., Frölich, J., & Lehmkuhl, G. (2013). Aufmerksamkeitsdefizit-/Hyperaktivitätsstörungen (ADHS) [Attention deficit/hyperactivity disorder]. In M. Döpfer, G. Lehmkuhl, & Petermann (Eds.), *Leitfaden Kinder- und Jugendpsychotherapie* (Bd. 1). Göttingen, Germany: Hogrefe.

Döpfner, M., Hautmann, C., Görtz-Dorten, A., Klasen, F., Ravens-Siebener, U., & The BELLA study group (2015). Long-term course of ADHD symptoms from childhood to early adulthood in a community sample. *European Child & Adolescent Psychiatry*, *24*, 665–673. doi:10.1007/s00787-014-0634-8.

Dumont, H., Maaz, K., Neumann, M., & Becker, M. (2014). Soziale Ungleichheiten im Übergang von Grundschule in die Sekundarstufe I: Theorie, Forschungsstand, Interventions- und Fördermöglichkeiten [Social disparities at the transition into secondary school: Theoretical conceptions and empirical evidence]. *Zeitschrift für Erziehungswissenschaften*, *17*, 141–165. doi:10.1007/s11618-013-0466-1.

EACEA (2016). Eurydice [Network]. Retrieved from https://webgate.ec.europa.eu/ fpfis/mwik is/eurydice/index.php/Main_Page.

Eells, W. C. (1930). Reliability of reported grading of examinations. *Journal of Educational Psychology*, *21*, 48–52.

Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb [Social background and competence acquisition]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 711–731). Münster, Germany: Waxmann.

Elliot, A. J. & McGregor, H. A. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*, 501–519. doi:10.1037/0022-3514.80.3.501.

Emmer, E. T. & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*, 103–112. doi:10.1207/S15326985EP3602_5.

Endepohls-Ulpe, M. & Ruf, H. (2005). Primary school teachers' criteria for the identification of gifted pupils. *High Ability Studies*, *16*, 219–228. doi:10.1080/13598130600618140.

Federal Statistical Office (2015). *Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus* [Population with migration background—results of micro-census]. Wiesbaden, Deutschland: Statistisches Bundesamt.

Feldhusen, J. F., Asher, J. W., & Hoover, S. M. (1984). Problems in the identification of giftedness, talent, or ability. *Gifted Child Quarterly*, *28*, 149–151. doi:10.1177/001698628402800402.

Feldhusen, J. F. & Moon, S. M. (1992). Grouping gifted students: Issues and concerns. *Gifted Child Quarterly*, *36*, 63–67. doi:10.1177/001698629203600202.

Fend, H. (1980). *Theorie der Schule* [Theory of school]. München, Germany: Urban & Schwarzenberg.

Fend, H. (2008). *Neue Theorie der Schule* [New theory of school]. Wiesbaden, Germany: VS Verlag.

Fennema, E., Peterson, P. L., Carpenter, T. P., & Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. *Educational Studies in Mathematics*, *21*, 55–65. doi:10.1007/bf00311015.

Fischbach, A. (2013). *Educational Assessment: Two sides of the same coin.* PhD thesis, University of Trier, Germany.

Fischbach, A., Baudson, T. G., Preckel, F., Martin, R., & Brunner, M. (2013). Do teacher judgments of student intelligence predict life outcomes? *Learning and Individual Differences*, *27*, 109–119. doi:10.1016/j.lindif.2013.07.004.

Fischer, C. & Müller, K. (2014). Gifted education and talent support in germany. *CEPS Journal*, *4*, 31–54.

Freudenthaler, H., Spinath, B., & Neubauer, A. (2008). Predicting school achievement in boys and girls. *European Journal of Personality*, *22*, 231–245. doi:10.1002/per.678.

Gangé, F. (1994). Are teachers really poor talent detectors? Comments on Pegnato and Birch's 1959 study of the effectiveness and efficacy of various identification techniques. *Gifted Child Quarterly*, *38*, 124–126. doi:10.1177/001698629403800305.

Gangé, F. (2004). Transforming gifts into talents: The DMGT as a developmental theory. *High Ability Studies*, *15*, 119–147. doi:10.1080/1359813042000314682.

García-Cepero, M. C. & McCoach, D. B. (2009). Educators' implicit theories of intelligence and beliefs about the identification of gifted students. *Universitas Psychologica*, *15*, 295–310.

Gear, G. H. (1979). Accuracy of teacher judgment in identifying intellectually gifted children: A review of the literature. *Gifted Child Quarterly*, *20*, 478–490. doi:10.1177/001698627602000416.

Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013). Prädiktoren der Schullaufbahnempfehlung für die Schulzweige des Sekundarbereichs I [Predictors of teacher track recommendation for secondary schooling]. *Pädagogische Rundschau*, *67*, 329–347.

Gniewosz, B. & Gräsel, C. (2011). Überblick Soziale Ungleichheit [Social disparities]. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Eds.), *Empirische Bildungsforschung – Gegenstandsbereiche* (pp. 163–167). Wiesbaden, Germany: VS Verlag.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23. doi:10.1016/S0160-2896(97)90011-8.

Gresch, C. & Becker, M. (2010). Sozial- und leistungsbedingte Disparitäten im Übergangsverhalten bei türkischstämmigen Kindern und Kindern aus (Spät-) Aussiedlerfamilien [Differences in secondary tracking related to social background and achievement of children with migration background]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 181–200). Bonn, Berlin, Germany: BMBF.

Gut, J., Reimann, G., & Grob, A. (2015). Kognitive, sprachliche, mathematische und sozial-emotionale Kompetenzen als Prädiktoren späterer schulischer Leistungen: Können die Leistungen eines Kindes in den IDS dessen Schulleistungen drei Jahre später vorhersagen? [Cognitive, language, mathematical, and socioemotional skills as predictors of subsequent school performance: Is a child's performance on the intelligence development scales predictive of its school performance three years later?]. *Zeitschrift für Pädagogische Psychologie*, *26*, 216–220. doi:10.1024/1010-0652/a000070.

Hanses, P. & Rost, D. H. (1998). Das "Drama" der hochbegabten Underachiever – "gewöhnliche" oder "außergewöhnliche" Under- achiever? [The "drama" of gifted underachievers—"ordinary" or "extraordinary" underachievers?]. *Zeitschrift für Pädagogische Psychologie*, *12*, 53–71.

Hany, E. A. (1997). Modeling teachers' judgment of giftedness: A methodological inquiry of biased judgment. *High Ability Studies*, *8*, 159–177. doi:10.1080/1359813970080203.

Hartog, P. & Rhodes, E. (1936). *An examination of examinations*. London, UK: McMillan.

Hell, B., Trapmann, S., Weigand, S., & Schuler, H. (2007a). Die Validität von Auswahlgesprächen im Rahmen der Hochschulzulassung – eine Metaanalyse [The validity of college admission interviews—A metaanalysis]. *Psychologische Rundschau*, *58*, 93–102.

Hell, B., Trapmann, S., Weigand, S., & Schuler, H. (2007b). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum [A meta-analytic investigation of subject-specific admission tests in German-speaking countries]. *Empirische Pädagogik*, *21*, 251–270.

Heller, K. A. (2004). Identification of gifted and talented students. *Psychology Science*, *46*, 302–323.

Heller, K. A. & Perleth, C. (2007). Talentförderung und Hochbegabtenberatung in Deutschand [Gifted education and counseling in germany]. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Eds.), *Begabt sein in Deutschland* (pp. 140–170). Berlin, Germany: LIT Verlag.

Helmke, A. (2007). *Unterrichtsqualität erfassen, bewerten, verbessern* [Assessment, evaluation, and improvement of teaching quality]. Stuttgart, Germany: Klett/Kallmeyer.

Helmke, A., Hosenfeld, I., & Schrader, F.-W. (2008). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften [Comparative class tests as an instrument to improve diagnostic competencies of teachers]. In R. Arnold & C. Griese

(Eds.), *Schulmanagement und Schulentwicklung* (pp. 113–135). Hohengehren, Germany: Schneider-Verlag.

Helmke, A., Rindermann, H., & Schrader, F.-W. (2008). Wirkfaktoren akademischer Leistungen in Schule und Hochschule [Influences on achievement in school and university]. In W. Schneider & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (pp. 145–155). Göttingen, Germany: Hogrefe.

Hernández-Torrano, D., Prieto, M. D., Ferrándiz, C., Bermejo, R., & Sáinz, M. (2013). Characteristics leading teachers to nominate secondary students as gifted in Spain. *Gifted Child Quarterly*, *57*, 181–196. doi:10.1177/0016986213490197.

Hesse, I. & Latzko, B. (2009). *Diagnostik für Lehrkräfte* [Diagnostics for teachers]. Opladen, Germany: Verlag Barbara Budrich.

Hillenbrand, C. & Ricking, H. (2011). Schulabbruch: Ursachen – Entwicklung – Prävention. Ergebnisse US-amerikanischer und deutscher Forschungen [Highschool dropout: Reasons—Developments—Prevention]. *Zeitschrift für Pädagogik*, *57*, 153–172.

Hinnant, J. B., O'Brien, N., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, *101*, 662–670. doi:10.1037/a0014306.

Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, *59*, 297–313. doi:10.3102/00346543059003297.

Holling, H. & Preckel, F. (2005). Self-estimates of intelligence – methodological approaches and gender differences. *Personality and Individual Differences*, *38*, 503–517. doi:10.1016/ j.paid.2004.05.003.

Hörstermann, T., Krolak-Schwerdt, S., & Fischbach, A. (2010). Die kognitive Repräsentation von Schülertypen bei angehenden Lehrkräften – Eine typologische Analyse [The cognitive representation of types of students in future teachers—A typological analysis]. *Schweizerische Zeitschrift für Bildungswissenschaften*, *32*, 143–158.

IBM Corp. (2012). IBM SPSS Statistics for Windows (Version 21.0) [statistical software]. Armonk, NY, USA: IBM Corp.

Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* [Textbook of pedagogical assessment]. Weinheim, Germany: Beltz.

Jacobs, C. & Petermann, F. (2007). Rechenstörungen [Dyscalculia]. In M. Döpfer, G. Lehmkuhl, & Petermann (Eds.), *Leitfaden Kinder- und Jugendpsychotherapie* (Bd. 9). Göttingen, Germany: Hogrefe.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*, 420–437.

Jonkmann, K., Maaz, K., Neumann, M., & Gresch, C. (2010). Übergangsquoten und Zusammenhänge zu familiärem Hintergrund und schulischen Leistungen: Deskriptive Befunde [Secondary school tracking and its relation to family background and school achievement]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 253–282). Bonn, Berlin, Germany: BMBF.

Jürges, H. & Schneider, K. (2008). Ressourcen und Anreize im Bildungswesen: Aufgaben und Handlungsmöglichkeiten des Staates aus Sicht der Bildungsökonomik [Ressources and incentives in education: Tasks and possibilities of state for action from the perspective of education economics]. *Zeitschrift für Erziehungswissenschaft*, *11*, 234–252. doi:10.1007/s11618-008-0024-4.

Jullien, S. (2006). *Elterliches Engagement und Lern- und Leistungsemotionen* [Parental involvement and achievement motivation]. Munich, Germany: Herbert Utz.

Jung, W. S. (2007). Preservice teacher training for successful inclusion. *Education*, *128*, 106–113.

Jussim, L. & Haber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*, 131–155. doi:10.1207/s15327957pspr0902_3.

Kaiser, J., Möller, J., Helm, F., & Kunter, M. (2015). Das Schülerinventar: Welche Schülermerkmale die Leistungsurteile von Lehrkräften beeinflussen [The student inventory: How student characteristics bias teacher judgments]. *Zeitschrift für Erziehungswissenschaft*, *18*, 279–302. doi:10.1007/s11618-015-0619-5.

Kantor, K. S. (2011). *General educators perceptions of preparedness to teach in mixed-ability classrooms*. PhD thesis, Walden University, Minneapolis, MN, USA.

Karg Foundation (2016). Fachportal Hochbegabung [Database Giftedness]. Retrieved from http://www.fachportal-hochbegabung.de/.

Karwowski, M. (2010). Are creative students really welcome in the classrooms? Implicit theories of "good" and "creative" student personality among Polish teachers. *Procedia Social and Behavioral Sciences*, *2*, 1233–1237. doi:10.1016/j.sbspro.2010.03.179.

Kaufman, J. C., Beghetto, R. A., & Dilley, A. (2016). Understanding creativity in schools. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century. Theory, research, and practice* (pp. 133–153). New York, NY, USA: Springer.

Kaufman, J. C., Plucker, J. A., & Russell, C. M. (2012). Identifying and assessing creativity as a component of giftedness. *Journal of Psychoeducational Assessment*, *30*, 60–73. doi:10.1177/0734282911428196.

Kelly, A. (1988). Gender differences in teacher-pupil interactions: A meta-analytic review. *Research in Education*, *39*, 1–23.

Klinge, D. (2016). *Die elterliche Übergangsentscheidung nach der Grundschule. Werte, Erwartungen und Orientierungen* [Parental enrollment decisions after primary school]. Wiesbaden, Germany: Springer.

Kloostermann, R., Ruiter, S., de Graf, P. M., & Kraaykamp, G. (2009). Parental education, children's performance and the transition to higher secondary education: trends in primary and secondary effects over five dutch school cohorts (1965–99). *British Journal of Sociology*, *60*, 377–398. British Journal of Sociology37-398.

Kohn, J., Wyschkon, A., Ballaschk, K., Ihle, W., & Esser, G. (2015). Verlauf von Umschriebenen Entwicklungsstörungen: Eine 30-Monats-Follow-up-Studie [Long-term course of dyslexia, dyscalculia and expressive language disorder: A 30-month follow-up study]. *Lernen und Lernstörungen*, *2*, 77–89. doi:10.1024/2235-0977/a000032.

Kristen, C. (2006). Ethnische Diskriminierung in der Grundschule? Die Vergabe von Noten und Bildungsempfehlungen [Ethnic discrimination in primary school?]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *58*, 79–97. doi:10.1007/s11575-006-0004-y.

Kucklick, C. (2011). Gute Lehrer [Quality teachers]. *GEO*, *02*, 24–48.

Kunter, M. & Klusmann, U. (2010). Die Suche nach dem kompetenten Lehrer – ein personenzentrierter Ansatz [Searching for the competent teacher]. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (pp. 207–230). Münster, Germany: Waxmann.

Legewie, J. & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *Sociological Review*, *77*, 463–485. doi:10.1177/0003122412440802.

Leutner, D. (2010). Pädagogisch-psychologische Diagnostik [pedago-psychological diagnostics]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (4rd ed.). (pp. 624–635). Weinheim, Germany: Beltz.

Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der schüler [The teacher is the key. Relation between techer competencies, teacher conduct and student learning]. In C. Allemann-Ghionda & E. Terhart (Eds.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern* (2nd ed.). (pp. 47–70). Oxford, UK: Pergamon.

Liu, J. (2004). Childhood externalizing behavior: Theory and implications. *Journal of Child and Adolescent Psychiatric Nursing*, *17*, 93–103. doi:10.1111/j.1744-6171.2004.tb00003.x.

Lohmann, H. & Groh-Samberg, O. (2010). Akzeptanz von Grundschulempfehlungen und Auswirkungen auf den weiteren Bildungsverlauf [Acceptance of secondary school track recommendations and their effects on educational achievement]. *Zeitschrift für Soziologie*, *39*, 470–492.

Loreman, T., Earle, C., Sharma, U., & Forlin, C. (2007). The development of an instrument for measuring pre-service teachers' sentiments, attitudes, and concerns about inclusive education. *International Journal of Special Education*, *22*, 150–159.

Lubinski, D. & Humphreys, L. G. (1990). A broadly based analysis of mathematical giftedness. *Intelligence*, *14*, 327–355. doi:10.1016/0160-2896(90)90022-l.

Luhmann, N. (2002). *Das Erziehungssystem der Gesellschaft* [Society's educational system]. Frankfurt am Main, Germany: Suhrkamp.

Maaz, K., Baumert, J., Gresch, C., & McElvany, N. (2010). *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* [The transistion from primary to secondary school—Educational justice and social disparities]. Bonn, Berlin, Germany: BMBF.

Maaz, K., Baumert, J., & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit?

[Emergence of social inequality in the institutional context of school: Where does social inequality emerge and grow?]. *Zeitschrift für Erziehungswissenschaft, Special Issue 12–09*, 11–46.

Maaz, K. & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte [The transition from elementary to secondary education in Germany: Definition, specification, and quantification of the primary and secondary effects of social background]. *Zeitschrift für Erziehungswissenschaft, Special Issue 12–09*, 153–182.

Mabe, P. & West, S. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280–296. doi:10.1037/0021-9010.67.3.280.

Machts, N. R., Kaiser, J., Schmidt, F., & Möller, J. (2016). Accuracy of teacher's judgments of students' cognitive abilities: A meta-analyses. *Educational Research Review, 19*, 85–103. doi:10.1016/j.edurev.2016.06.003.

Macnamara, J. & Madaus, G. F. (1969). Marker reliability in the irish leaving certificate. *Irish Journal of Education, 3*, 5–21.

Madelaine, A. & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education, 52*, 33–43. doi:10.1080/10349120500071886.

Marland, S. P. (1972). *Education of the gifted and talented*. Retrieved from https://www.valdosta.edu/colleges/education/psychology-and-counseling/documents/marland-report.pdf.

Marsh, H. W. (1984). Relations among dimensions of self-attribution, dimensions of self-concept, and academic achievements. *Journal of Educational Psychology, 76*, 1291–1308.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, , 280–295.

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*, 510–524. doi:10.1037/0022-0663.100.3.510.

Marwege, G. (2013). *Legasthenie und Dyskalkulie in der Schule* [Dyslexia and dyscalculia in the school]. Göttingen, Germany: Universitätsverlag Göttingen.

McBee, M. (2010). Examining the probability of identification for gifted programs for students in georgia elementary schools: A multilevel path analysis study. *Gifted Child Quarterly*, *54*, 283–297. doi:10.1177/0016986210377927.

McClain, M.-C. & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology*, *28*, 59–88. doi:10.1080/15377903.2012.643757.

McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern [Teachers' diagnostic skills to judge student performance and task difficulty when learning materials include instructional pictures]. *Zeitschrift für Pädagogische Psychologie*, *23*, 223–235. doi:10.1024/1010-0652.23.34.223.

Meier, E., Vogl, K., & Preckel, F. (2014). Motivational characteristics of students in gifted classes: The pivotal role of need for cognition. *Learning and Individual Differences*, *33*, 39–46. doi:10.1016/j.lindif.2014.04.006.

Meulemann, H. & Relikowski, I. (2016). Chancengleichheit und Leistung im sozialen Wandel. Primäre und sekundäre Effekte der sozialen Herkunft am Gymnasialübergang in Hessen 1969 und 2007 [Equality of opportunity and achievement in social change. Primary and secondary effects of social origin at secondary school transition in Hesse 1969 and 2007]. *Zeitschrift für Erziehungswissenschaften*, *19*, 443–462. doi:10.1007/s11618-015-0654-2.

Mönks, F. J. & Heller, K. A. (1994). Identification and programming of the gifted and talented. In T. Husén & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed.). (pp. 2725–2732). Oxford, UK: Pergamon.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 International Report*. Chestnut Hill, MA, USA: TIMSS & PIRLS International Study Center.

Mullola, S., Ravaj, N., Lipsanen, J., Alatupa, S., Hintsanen, M., Jokela, M., & Keltikangas-Järvinen, L. (2012). Gender differences in teachers' perceptions of students' tempera-

ment, educational competence, and teachability. *British Journal of Educational Psychology*, *82*, 185–206. doi:10.1111/j.2044-8279.2010.02017.x.

Muthén, L. K. & Muthén, B. O. (1998–2015). Mplus Version 7.31 [statistical software]. Los Angeles, CA, USA: Muthén & Muthén.

Neber, H. (1994). Teacher identification of students for gifted programs: Nomination to a summer school for highly-gifted students. *Psychology Science*, *46*, 348–362.

Neisser, U., Boodoo, G., Jr., T. L. B., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sernberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101. doi:10.1037//0003-066X.51.2.77.

Neugebauer, M. (2010). Bildungsungleichheit auf Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte [Inequality and teacher recommendations at the transition to upper secondary school: A decomposition of primary and secondary effects of social origin]. *Zeitschrift für Soziologie*, *39*, 202–214.

Neugebauer, M. (2011). Werden Jungen von Lehrerinnen bei den Übergangsempfehlungen für das Gymnasium benachteiligt? Eine Analyse auf Basis der IGLU-Daten [Are boys disadvantaged when female teachers resommend a secondary school track?]. In A. Hadjar (Ed.), *Geschlechtsspezifische Bildungsungleichheiten* (pp. 235–260). Wiesbaden, Germany: Springer.

Neumann, M., Maaz, K., & Becker, M. (2013). Die Abkehr von der traditionellen Dreigliedrigkeit im Sekundarschulsystem: Auf unterschiedlichen Wegen zum gleichen Ziel? [The development from a three-tier to a two-tier secondary school system]. *Recht der Jugend und des Bildungswesens*, *61*, 274–292.

Neumann, M., Milek, A., Maaz, K., & Gresch, C. (2010). Zum Einfluss der Klassenzusammensetzung auf den Übergang von der Grundschule in die weiterführenden Schulen [The influence of class characteristics on the transition from primary to secondary school]. In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule - Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 229–251). Bonn, Berlin, Germany: BMBF.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159. doi:10.1037/a0026699.

OECD (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris, France: OECD.

OECD (2011). *PISA 2009 Ergebnisse: Potenziale nutzen und Chancengleichheit sichern – Sozialer Hintergrund und Schülerleistungen* [PISA 2009 Results: Overcoming Social Back-ground—Equity in Learning Opportunities and Outcomes], volume 2. OECD. doi:10.1787/ 9789264095359-de.

Petersen, J. (2013). Gender differences in the identification of gifted youth and in gifted program participation: A meta-analysis. *Contemporary Educational Psychology*, *38*, 342–348. doi:10.1016/j.cedpsych.2013.07.002.

Pfeiffer, S. I. (2012). Current perspectives on the identification and assessment of gifted students. *Journal of Psychoeducational Assessment*, *30*, 3–9. doi:10.1177/ 0734282911428192.

Pfeiffer, S. I. & Jarosewich, T. (2003). *Gifted Rating Scales—School form*. San Antonio, TX, USA: Harcourt Assessment.

Plake, B. S. & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 53–68). San Diego, CA, USA: Academic Press.

Plucker, J., Beghetto, R. A., & Dow, G. (2004). Why isn't creativity more important to educational psychologists? potential, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*, 83–96. doi:10.1207/s15326985ep3902_1.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*, 4–11. doi:10.1080/00405840802577536.

Praetorius, A.-K., Karst, K., Dickhäuser, O., & Lipowsky, F. (2011). Wie gut schätzen Lehrer die Fähigkeitsselbstkonzepte ihrer Schüler ein? Zur diagnostischen Kompetenz von lehrkräfte [How accurate do teachers judge their students ability self-concepts? Regarding teachers' diagnostic competence]. *Psychologie in Erziehung und Unterricht*, *58*, 81–91. doi:10.2378/peu2011.art30d.

Preckel, F. (2008). Erkennen und Fördern hochbegabter Schülerinnen und Schüler [Identifying and educating gifted students]. In W. Schneider & F. Petermann (Eds.), *Angewandte Entwicklungspsychologie. Enzyklopädie der Psychologie* (pp. 449–495). Göttingen, Germany: Hogrefe.

Preckel, F., Baudson, T. G., Krolak-Schwert, S., & Glock, S. (2015). Gifted and maladjusted? implicit attitudes and automatic associations related to gifted children. *American Educational Research Journal*, *52*, 1160–1184. doi:10.3102/0002831215596413.

Preckel, F., Götz, T., & Frenzel, A. C. (2010). Ability grouping of gifted students: Effects on academic self-concept and boredom. *British Journal of Educational Psychology*, *80*, 451–472. doi:10.1348/000709909X480716.

Prenzel, M., Sälzer, C., Klieme, E., & Köller, O. (2013). *PISA 2012 – Fortschritte und Herausforderungen in Deutschland* [PISA 2012—Progress and challenges in Germany]. Münster, Germany: Waxmann.

Rauer, W. & Schuck, K. D. (2003). *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen – FEESS 3–4* [Social and emotional school experiences survey—Grades 3 and 4]. Göttingen, Germany: Hogrefe.

Rauer, W. & Schuck, K. D. (2004). *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen – FEESS 1–2* [Social and emotional school experiences survey—Grades 1 and 2]. Göttingen, Germany: Hogrefe.

Reis, S. M. & McCoach, D. B. (1986). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly*, *44*, 152–170. doi:10.1177/001698620004400302.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *The Phi Delta Kappan*, *60*, 180–184. doi:10.1177/003172171109200821.

Renzulli, J. S. (2005). The three-ring conception of giftedness: A developmental model for promoting creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed.). (pp. 246–279). Cambridge, NY, USA: Cambridge University Press.

Renzulli, J. S. & Delcourt, M. A. B. (1986). The legacy and logic of research on the identification of gifted persons. *Gifted Child Quarterly*, *30*, 20–23. doi:10.1177/001698628603000104.

Renzulli, J. S. & Reis, S. M. (1985). *The schoolwide enrichment model: A comprehensive plan for educational excellence*. Mansfield Center, CT, USA: Creative Learning Press.

Rimm, S. B. (2003). Underachievement: A national epidemic. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed.). (pp. 424–443). Boston, MA, USA: Allyn & Bacon.

Rischke, M., Baedorf, D., & Müller, U. (2015). Inklusionsorientierte Lehrerbildung – vom Schlagwort zur Realität?! [Inclusive teacher education—from keyword to reality?!]. *Monitor Lehrerbildung*, *05*.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, *73*, 417–458.

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly*, *51*, 382–396. doi:10.1177/0016986207306324.

Rohrer, J. C. (1995). Primary teacher [sic!]conception of giftedness: Image, evidence, and nonevidence. *Journal for the Education of the Gifted*, *18*, 269–283.

Roos, J., Schöler, H., Zöller, I., & Treutlein, A. (2011). PRISE – Der Übergang vom Primar- zum Sekundarbereich. Einflüsse der Entwicklung im Grundschulalter auf die Entwicklung in den ersten zwei Jahren der Sekundarschulzeit [The transition from primary to secondary school]. In Baden-Württemberg Stiftung gGmbH (Ed.), *Programm Bildungsforschung – Projektergebnisse* (pp. 55–80). Stuttbarg, Germany: Baden-Württemberg Stiftung gGmbH.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. doi: 10.1016/j.intell.2015.09.002.

Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology*, -. doi:10.1037/edu0000085.

Roznowski, M., Hong, S., & Reith, J. (2000). A further look at youth intellectual giftedness and its correlates: Values, interests, performance, and behavior. *Intelligence*, *28*, 87–113. doi:10.1016/S0160-2896(99)00032-X.

Sacher, W. (2009). *Leistungen entwickeln, überprüfen und beurteilen* [Development, assessment, and grading of performance]. Bad Heilbrunn, Germany: Julius Klinkhardt.

Sahlberg, P. (2007). Education policies for raising student learning: The Finnish approach. *Journal of Education Policy*, *22*, 173–197.

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory Into Practice*, *32*, 118–126. doi:10.1080/00405849309543585.

Scharenberg, K., Gröhlich, C., Guill, K., & Bos, W. (2010). Schulformwechsel und prognostische Validität der Schullaufbahnempfehlung in der Jahrgangsstufe 4 [Changes in secondary school tracks and prognostic validity of track recommendations at the end of forth grade]. In W. Bos & C. Gröhlich (Eds.), *HANSE – Hamburger Schriften zur Qualität im Bildungswesen* (pp. 119–141). Münster, Germany: Waxmann.

Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung [The importance of social and migrant backgrounds for school recommendations in Germany at the end of primary school]. *Zeitschrift für Erziehungswissenschaft*, *14*, 371–396. doi:10.1007/s11618-011-0221-4.

Schrader, F.-W. (2006). Diagnostische Kompetenz von Eltern und Lehrern [Teachers' and parents' diagnostic abilities]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (3rd ed.). (pp. 95–100). Weinheim, Germany: Beltz.

Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen [Teachers' diagnostic competency: Components and effects]. *Empirische Pädagogik*, *1*, 27–52.

Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer [Everyday classroom assessment by teachers]. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (pp. 45–58). Weinheim, Germany: Beltz.

Schründer-Lenzen, A. & Merkens, H. (2006). Differenzen schriftsprachlicher Kompetenzentwicklung bei Kindern mit und ohne Migrationshintergrund [Differences in writing literacy between children with and without migration background]. In A. Schründer-Lenzen (Ed.), *Risikofaktoren kindlicher Entwicklung* (pp. 15–44). Wiesbaden, Germany: VS Verlag.

Schuchart, C. & Weihaupt, H. (2004). Die prognostische Qualität der Übergangsempfehlungen der niedersächsischen Orientierungsstufe [The prognostic quality of secondary track recommendations]. *Zeitschrift für Pädagogik*, *50*, 882–902.

Scott, C. (1999). Teachers' biases toward creative children. *Creativity Research Journal*, *12*, 321–328. doi:10.1207/s15326934crj1204_10.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*, 743–762. doi:10.1037/a0027627.

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum: Eine experimentelle Untersuchung zur diagnostischen Kompetenz [The simulated classroom: An experimental study on diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, *22*, 261–276. doi:10.1024/1010-0652.22.34.261.

Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 275–326). San Francisco, CA, USA: Wiley.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*, 4–14. doi:10.3102/0013189X015002004.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 1–22. doi:10.17763/haer.57.1.j463w79r56455411.

Shulman, L. S. (1998). Theory, practice, and the education of professionals. *Elementary School Journal*, *98*, 511–526. doi:10.17763/haer.57.1.j463w79r56455411.

Siegle, D. & Powell, T. (2004). Exploring teacher biases when nominating students for gifted programs. *Gifted Child Quarterly*, *48*, 21–29. doi:10.1177/001698620404800103.

Siegle, D. & Reis, S. M. (1998). Gender differences in teacher and student perceptions of gifted students' ability. *Gifted Child Quarterly*, *42*, 39–48. doi:10.1177/001698629804200105.

Süßlin, W. (2012). Die Situation an den deutschen Schulen aus Sicht von Lehrern und Eltern [The German school situation from teachers' and parents' perspectives]. In Vodafone Stiftung Deutschland (Ed.), *Lehre(r) in Zeiten der Bildungspanik. Eine Studie zum Prestige des Lehrerberufs und zur Situation an den Schulen in Deutschland* (pp. 6–46). Düsseldorf, Germany: Vodafone Stiftung Deutschland.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on

student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie*, *19*, 85–95. doi:10.1024/1010-0652.19.12.85.

Spinath, B., Freudenthaler, H. H., & Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences*, *48*, 481–486. doi:10.1016/j.paid.2009.11.028.

Spniath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, *34*, 363–374. doi:10.1016/j.intell.2005.11.004.

Stanat, P. & Christensen, G. (2006). *Schulerfolg von Jugendlichen mit Migrationshintergrund im internationalen Vergleich. Eine Analyse von Voraussetzungen und Erträgen schulischen Lernens im Rahmen von PISA 2003* [School success of teenagers with migration backgrounds in international comparison]. Bonn, Berlin: Germany: BMBF.

Stanat, P., Rauch, D., & Segeritz, M. (2010). PISA 2009. Bilanz nach einem Jahrzehnt [Students wirh migration background]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt* (pp. 200–230). Münster, Germany: Waxmann.

Stanat, P., Segeritz, M., & Christensen, G. (2010). Schulbezogene Motivation und Aspiration von Schülerinnen und Schülern mit Migrationshintergrund [School-realted motivation and aspirations of students with migration background]. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (pp. 31–58). Münster, Germany: Waxmann.

Stang, J. & Urhahne, D. (2016). Wie gut schätzen Lehrkräfte Leistung, Konzentration, Arbeits- und Sozialverhalten ihrer Schülerinnen und Schüler ein? Ein Beitrag zur diagnostischen Kompetenz von Lehrkräften [How teachers rate students achievement, attention, work habits and social behavior?]. *Psychologie in Erziehung und Unterricht*, *63*, 204–219. doi:10.2378/peu2016.art18d.

Starch, D. & Elliot, E. C. (1912). Reliability of grading high school work in english. *School Review*, *20*, 442–457.

Starch, D. & Elliot, E. C. (1913). Reliability of grading high school work in mathematics. *School Review*, *21*, 254–259.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York, NY, USA: Cambridge University Press.

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, *72*, 2534–539.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, *64*, 22–26.

Stubbe, T. C. & Bos, W. (2008). Schullaufbahnempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangstufe [Secondary track recommendations and eonrollment decisions at the end of fourth grade]. *Empirische Pädagogik*, *50*, 882–902.

Tarelli, I., Valtin, R., Bos, W., Bremerich-Vos, A., & Schwippert, K. (2012). IGLU 2011: Wichtige Ergebnisse im Überblick [ILGU 2011: Important findings]. In W. Bos, I. Tarelli, A. Bremerich-Vos, & K. Schwippert (Eds.), *IGLU 2011 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 11–25). Münster, Germany: Waxmann.

Tenenbaum, H. R. & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for european american students? A meta-analysis. *Journal of Educational Psychology*, *99*, 253–273. doi:10.1037/0022-0663.99.2.253.

Tent, L. & Stelzl, I. (1993). *Pädagogisch-psychologische Diagnostik: Theoretische und methodische Grundlagen* [Pedago-psychological diagnostics: Theoretical and methodological basics]. Göttingen, Germany: Hogrefe.

The United Nations (1948). *Universal Declaration of Human Rights*. UN.

Trapmann, S., Hell, B., Weigand, S., & Schuler, H. S. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine metaanalyse [The validity of school grades for academic achievement—A meta-analysis]. *Zeitschrift für Pädagogische Psychologie*, *21*, 11–27. doi:10.1024/1010-0652.21.1.11.

Trautwein, U. (2014). Wie stark trägt das gegliederte Schulsystem zur sozialen Ungleichheit bei? [How does the German tracked secondary school system increase social disparities?]. In B. Spinath (Ed.), *Empirische Bildungsforschung. Aktuelle Themen der Bildungspraxis und Bildungsforschung* (pp. 68–78). Berlin Heidelberg, Germany: Springer.

Trautwein, U. & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übertrittsentscheidungen [When high-achie-

ving classmates put students at a disadvantage: Reference group effects at the transition to secondary schooling]. *Zeitschrift für Pädagogische Psychologie*, *21*, 119–133. doi:10.1024/1010-0652.21.2.119.

Urhahne, D. & Zhu, M. (2015). Teacher judgement and student motivation. In C. M. Rubie-Davies, J. M. Stephens, & P. Watson (Eds.), *International Handbook of Social Psychology of the Classroom* (pp. 304–315). London, UK: Routledge.

Valdez, A. (2013). Teacher judgement of reading achievement: Cross-sectional and longitudinal perspective. *Journal of Education and Learning*, *2*, 186–200. doi:10.5539/jel.v2n4p186.

Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell [Relations of general and specific intelligence factors with general and specific achievement factors in a nested-factor-model]. *Zeitschrift für Pädagogische Psychologie*, *29*, 101–108. doi:10.1024/1010-0652/a000151.

van Yperen, N. W. (2003). Task interest and actual performance: The moderating effects of assigned and adopted purpose. *Journal of Personality and Social Psychology*, *85*, 1006–1015. doi:10.1037/0022-3514.85.6.1006.

Välijärvi, J. & Malin, A. (2003). The two-level effect of socio-economic background. In S. Lie, P. Linnakylä, & A. Roe (Eds.), *Northern lights on PISA: unity and diversity in Nordic countries in PISA 2000* (pp. 132–132). Oslo, Norway: University of Oslo.

Vock, M., Preckel, F., & Holling, H. (2011). Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence*, *39*, 357–369. doi:10.1016/j.intell.2011.06.006.

Voss, T. (2010). *Generelles pädagogisch-psychologisches Wissen und lerntheoretische Überzeugungen von Lehrkräften als Aspekte der Lehrerkompetenz* [Teachers' general pedagogical/psychological knowledge and beliefs as aspects of their professional competence]. PhD thesis, Freie Universität Berlin, Germany.

Voss, T., Kunina-Habenicht, O., Hoehne, V., & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde [teachers' pedagogical knowledge: Empirical approaches and findings]. *Zeitschrift für Erziehungswissenschaft*, *18*, 187–223. doi:10.1007/s11618-015-0626-6.

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical and psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, *103*, 952–969.

Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Journal of Education and Learning*, *140*, 1174–1204. doi:10.1037/a0036620.

Warnke, A., Hemminger, U., & Plume, E. (2004). Lese-Rechtschreibstörungen [Dyslexia]. In M. Döpfer, G. Lehmkuhl, & Petermann (Eds.), *Leitfaden Kinder- und Jugendpsychotherapie* (Bd. 6). Göttingen, Germany: Hogrefe.

Weinert, F. E. (2001). Schulleistungen – Leistungen der Schule oder der Schüler [School achievement—achievement of the schools or the students]. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (pp. 68–78). Weinheim, Germany: Beltz.

Wentzel, K. R. (1993). Does being good make the grade? social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*, 357–364. doi:10.1037/0022-0663.85.2.357.

Westbury, I., Hansen, S.-E., Kansanen, P., & Björkvist, O. (2005). Teacher education for research-based practice in expanded roles: Finland's experience. *Scandinavian Journal of Educational Research*, *49*, 475–485. doi:10.1080/00313830500267937.

Westby, E. L. & Dawson, V. L. (1995). Creativity: Asset or burden in the classroom? *Creativity Research Journal*, *8*, 1–10. doi:10.1207/s15326934crj0801_1.

Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC, USA: American Psychological Association.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York, NY, USA: McGraw-Hill.

Wolf, K. P. (1993). From informal to informed assessment: Recognizing the role of the classroom teacher. *Journal of Reading*, *36*, 518–523.

Wollschläger, R., Baudson, T. G., Schmitt, J., Fendahl, S., & Preckel, F. (2014). *Referenzgruppeneffekte der mittleren Intelligenz einer Klasse auf Fähigkeitsbeurteilungen durch Lehrkräfte* [Reference-group-effects of class mean intelligence on teacher judgments].

Vortrag bei der 2. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF, 03.–05. März 2014), Frankfurt am Main, Germany.

Wu, A. D., Liu, Y., Gadermann, A. M., & Zumbo, B. D. (2010). Multiple-indicator multilevel growth model: A solution to multiple methodological challenges in longitudinal studies. *Social Indicators Research: International Interdisciplinary Journal for Quality of Life Measurement*, *97*, 123–142. doi:10.1007/s11205-009-9496-8.

Wyatt-Smit, C., Castleton, G., Freebody, P., & Cooksey, R. W. (2003). The nature of teachers' qualitative judgements: A matter of context and salience. *Australian Journal of Language and Literacy*, *26*, 11–32. doi:pes:1093.

A. Appendix

A.1. Tables

A.1.1. Preliminary analyses of Study 1

Table A.1.: Results of measurement invariance analyses of teacher judgment of cognitive ability over Grades

	χ^2	*df*	**Scaling correction factor**	$\Delta\chi^2$ (*df*), *p*	**CFI**	**ΔCFI**	**RMSEA** [**90%CI**]
Sample 1							
Configural	0.000	0	1.0000		1.00		.000 [.000 − .000]
Metric	6.165	2	1.5535	6.165 (2), .046	.995	−.005	.079 [.009 − .152]
Scalar	7.283	4	1.5930	7.283 (4), .121	.996	+.001	.049 [.000 − .105]
Strict	19.606	7	1.7731	11.505 (3), .009	.986	−.010	.037 [.036 − .112]
Sample 2							
Configural	0.000	0	1.0000		1.00		.000 [.000 − .000]
Metric	13.340	6	1.4162	13.340 (6), .038	.997	−.003	.053 [.012 − .092]
Scalar	22.587	12	1.6697	22.587 (12), .031	.995	−.002	.045 [.013 − .073]
Strict	25.518	21	2.1062	6.007 (9), .739	.998	+.002	.022 [.000 − .048]

Note. Configural measurement could not be tested because the model was completely identified. Goodness-of-fit indices of models tested in each group separately were thus not reported, since they were as those depicted in the configural measurement invariance test. χ^2 difference tests between scalar and strict measurement invariance models were conducted with the adjusted χ^2 using the Satorra-Bentler scaling correction due to the MLR estimation in M*plus* 7.31; all other χ^2 difference tests were run within M*plus* 7.31 automatically.

A.1.2. Preliminary analyses of Study 2

Table A.2.: Descriptive statistics, reliability, and Intra-Class-Correlation of IQ, teacher judgments, and probability of nomination by Grade

	N	*M (SD)*	*Min.*	*Max.*	Cronbach's α	ICC
Grade 2						
IQ (THINK)	679	100.81 (14.10)	70.55	140.30		
TJ_cognitive ability	677	3.42 (1.31)	.00	6.00	.96	.002
TJ_academic motivation	658	3.24 (.96)	.00	6.00	.66	.057
TJ_creativity	663	3.26 (1.14)	.00	6.00	.91	.025
TJ_participation in class	671	3.51 (1.21)	.33	6.00	.85	.022
TJ_probability of nomination	679	32.85 (28.50)	.00	100.00		.243
Grade 3						
IQ (THINK)	679	100.81 (14.10)	70.55	140.30		
TJ_cognitive ability	677	3.42 (1.31)	.00	6.00	.96	.002
TJ_academic motivation	658	3.24 (.96)	.00	6.00	.66	.057
TJ_creativity	663	3.26 (1.14)	.00	6.00	.91	.025
TJ_participation in class	671	3.51 (1.21)	.33	6.00	.85	.022
TJ_probability of nomination	679	32.85 (28.50)	.00	100.00		.243

Note. TJ_ = teacher judgment.

Table A.3.: Results of confirmatory factor analyses of the teacher questionnaire

Model	$\chi^2(df)$	χ^2/df	**CFI**	**RMSEA** [**90%CI**]	**SRMR**
g-Factor model	1008.11 (54)***	18.67	.784	.161 [.153 – .170]	.079
2-Factor model	683.76 (53)***	12.90	.857	.132 [.124 – .141]	.059
3-Factor model	525.57 (51)***	10.31	.892	.117 [.108 – .126]	.048
4-Factor model	289.28 (48)***	6.03	.945	.086 [.077 – .096]	.040

Note. The models were specified as follows: *g*-Factor model: all items loading on one superordinate factor. 2-Factor model: TJs of cognitive ability and creativity were collapsed into one factor, academic motivation and participation in class into another. 3-Factor model: TJs of cognitive ability and creativity were collapsed into one factor, academic motivation and participation in class were treated as separate factors. 4-Factor model: TJs of all four constructs were modeled as separate factors. Factor intercorrelations were allowed in all multi-factor models.

*** $p < .001$

Table A.4.: Results of measurement invariance analyses of teacher judgments over grades

	χ^2	*df*	**Scaling correction factor**	$\Delta\chi^2$ $(df), p$	**CFI**	**ΔCFI**	**RMSEA** [**90%CI**]
Teacher Judgment of Cognitive Ability							
Configural	0.00	0	1.00		1.00		.000 [.000 – .000]
Metric	6.17	2	1.55	6.17 (2), .046	.995	−.005	.079 [.009 – .152]
Scalar	7.28	4	1.59	7.28 (4), .121	.996	+.001	.049 [.000 – .105]
Strict	19.61	7	1.77	11.51 (3), .009	.986	−.010	.037 [.036 – .112]
Teacher Judgment of Creativity							
Configural	0.00	0	1.00		1.00		.000 [.000 – .000]
Metric	0.53	2	1.46	0.53 (2), .768	1.00	−.000	.000 [.000 – .072]
Scalar	2.78	4	1.00	2.78 (4), .596	1.00	−.000	.000 [.000 – .069]
Strict	12.03	7	1.52	7.01 (3), .072	.994	−.006	.046 [.000 – .089]
Teacher Judgment of Academic Motivation (partial measurement invariance)							
Configural	0.00	1	1.50		1.00		.000 [.000 – .000]
Metric	1.52	3	1.89	1.37 (2), .504	1.00	−.000	.000 [.000 – .070]
Scalar	1.90	5	2.03	1.78 (4), .776	1.00	−.000	.000 [.000 – .040]
Strict	2.80	7	2.25	0.49 (2), .784	1.00	−.006	.000 [.000 – .028]
Teacher Judgment of Participation in Class							
Configural	0.00	0	1.00		1.00		.000 [.000 – .000]
Metric	0.44	2	1.58	0.44 (2), .805	1.00	−.000	.000 [.000 – .067]
Scalar	4.59	4	1.61	4.59 (4), .332	.999	−.001	.021 [.000 – .087]
Strict	6.59	7	1.59	1.97 (3), .579	1.00	+.001	.000 [.000 – .066]

Note. Configural measurement could not be tested because the model was completely identified. Goodness-of-fit indices of models tested in each group separately were thus not reported, since they were as those depicted in the configural measurement invariance test. χ^2 difference tests between scalar and strict measurement invariance models were conducted with the adjusted χ^2 using the Satorra-Bentler scaling correction due to the MLR estimation in M*plus* 7.31; all other χ^2 difference tests were run within M*plus* 7.31 automatically.

Table A.5.: Results of measurement invariance analyses of teacher judgments over gender

	χ^2	*df*	**Scaling correction factor**	$\Delta\chi^2$ $(df), p$	**CFI**	**ΔCFI**	**RMSEA** [**90%CI**]
Teacher Judgment of Cognitive Ability							
Configural	0.00	0	1.00		1.00		.000 [.000 – .000]
Metric	2.86	2	1.04	2.86 (2), .240	.999	−.001	.036 [.000 – .120]
Scalar	5.10	4	1.10	5.10 (4), .277	.999	−.000	.029 [.000 – .091]
Strict	5.53	7	1.44	1.24 (3), .743	1.00	+.001	.000 [.000 – .580]
Teacher Judgment of Creativity							
Configural	0.00	0	1.00		1.00		.000 [.000 – .000]
Metric	0.88	2	1.16	0.88 (2), .645	1.00	−.000	.000 [.000 – .085]
Scalar	6.88	4	1.50	6.88 (4), .142	.996	−.004	.000 [.000 – .103]
Strict	19.36	7	1.31	14.23 (3), .003	.985	−.011	.072 [.035 – .112]
Teacher Judgment of Academic Motivation (partial measurement invariance)							
Configural	0.21	1	1.15		1.00		.000 [.000 – .011]
Metric	3.38	3	1.53	2.87 (2), .239	.999	−.001	.019 [.000 – .096]
Scalar	22.71	5	1.50	21.32 (4), .000	.937	−.062	.102 [.062 – .147]
Teacher Judgment of Participation in Class (partial measurement invariance)							
Configural	1.74	1	1.37		.999		.047 [.000 – .061]
Metric	3.69	3	1.67	2.08 (2), .353	.999	−.000	.026 [.000 – .099]
Scalar	7.90	5	1.55	6.18 (4), .186	.995	−.004	.041 [.000 – .093]
Strict	9.53	7	1.53	1.58 (2), .454	.996	+.001	.033 [.000 – .079]

Note. Configural measurement could not be tested because the model was completely identified. Goodness-of-fit indices of models tested in each group separately were thus not reported, since they were as those depicted in the configural measurement invariance test. χ^2 difference tests between scalar and strict measurement invariance models were conducted with the adjusted χ^2 using the Satorra-Bentler scaling correction due to the MLR estimation in M*plus* 7.31; all other χ^2 difference tests were run within M*plus* 7.31 automatically.

Authorship and publication status

Chapters 4 to 6 present research articles that are or will be submitted for publication. Hereinafter the authors and publications statuses of the papers are enlisted.

1. Study 1 (Chapter 4): Manuscript soon to be submitted for publication in *Journal of Educational Research*: Wollschläger., R., Baudson, T. G., & Preckel, F. (2016). Accuracy of teacher judgments of primary school students' cognitive ability.

2. Study 2 (Chapter 5): Manuscript submitted for publication and currently under review in *Gifted Child Quarterly*: Wollschläger., R., Baudson, T. G., & Preckel, F. (2016). Keeping the gate: Components of primary school teachers' nominations for a gifted program in Germany.

3. Study 3 (Chapter 6): Manuscript soon to be submitted for publication in *Zeitschrift für Pädagogische Psychologie*: Wollschläger., R. & Preckel, F. (2016). Teachers' secondary school track recommendations and parental enrollment decisions against students' social background.

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebene Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quellen gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere die Richtigkeit der vorangegangenen Erklärung und bin mir der strafrechtlichen Folgen einer Falschaussage bewusst.

Trier, 20. September 2016

Dipl.-Psych. Rachel Wollschläger