

Knigge, Michel; Nordstrand, Vibeke; Walzebug, Anke
Do teacher stereotypes about school tracks function as expectations at the collective level and do they relate to the perception of obstacles in the classroom and to teachers' self-efficacy beliefs?

Journal for educational research online 8 (2016) 2, S. 158-191



Empfohlene Zitierung/ Suggested Citation:

Knigge, Michel; Nordstrand, Vibeke; Walzebug, Anke: Do teacher stereotypes about school tracks function as expectations at the collective level and do they relate to the perception of obstacles in the classroom and to teachers' self-efficacy beliefs? - In: Journal for educational research online 8 (2016) 2, S. 158-191 - URN: urn:nbn:de:0111-pedocs-124406

in Kooperation mit / in cooperation with:

WAXMANN
VERLAG GMBH
Münster · New York · München · Berlin



<http://www.waxmann.com>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Michel Knigge, Vibeke Nordstrand & Anke Walzebug

Do teacher stereotypes about school tracks function as expectations at the collective level and do they relate to the perception of obstacles in the classroom and to teachers' self-efficacy beliefs?

Abstract

The effects of individual teacher expectations have been the subject of intensive research. Results indicate that teachers use their expectations to adapt their interactions with their students to some degree (as summarized in a review by Jussim & Harber, 2005). This can in turn lead to expectancy-confirming student developments. While there are studies on the Pygmalion effect on individual students, there is only little research on teacher judgements of whole classes and schools. Our study aims to extend the perspective of teacher judgements at the collective level to stereotypes within the context of school tracking. The content and structure of teachers' school track stereotypes are investigated as well as the question of whether these stereotypical judgements are related to teachers' perception of obstacles to their teaching and their teaching self-efficacy beliefs. Cross-sectional data on 341 teachers at two different school types from the Panel Study at the Research School „Education and Capabilities“ in North Rhine-Westphalia (PARS) (see Bos et al., 2016) were used for two purposes: First, the structure of teachers' stereotypes was identified via an exploratory factor analysis. Second, in follow-up regression analyses, the stereotype dimensions extracted were used to predict teachers' perceptions of obstacles to their classroom work and their individual and collective teacher self-efficacy beliefs. Results showed that – after controlling for the average cognitive abilities and the average cultural capital of the students

Prof. Dr. Michel Knigge (corresponding author), Human Science Faculty, Educational Science, University of Potsdam, Karl-Liebknecht-Straße 24–25, Haus 31, 14476 Potsdam, Germany

e-mail: michel.knigge@uni-potsdam.de

Vibeke Nordstrand, Institute for Educational Science, University of Hamburg, Von-Melle-Park 8, 20146 Hamburg, Germany

e-mail: vibeke.nordstrand@uni-hamburg.de

Dr. Anke Walzebug, Institute for Education Research and School Development (IFS), TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany

e-mail: anke.walzebug@tu-dortmund.de

– teacher stereotypes were indeed related to perceived obstacles concerning their classroom work and their self-efficacy beliefs. After a discussion of the strengths and limitations of the present research, the article closes with a short proposal of a future research framework for collective Pygmalion effects.

Keywords

Teacher; Stereotypes; Self-Efficacy; Pygmalion; Collective expectations

Stereotypen über Schulformen als Erwartungen von Lehrkräften auf der kollektiven Ebene – Zusammenhänge mit Selbstwirksamkeitserwartungen und wahrgenommenen Schwierigkeiten im Klassenzimmer

Zusammenfassung

Erwartungseffekte auf der Ebene von Einzelschülerinnen und Einzelschülern sind ein gut untersuchtes Phänomen. Im Gesamtergebnis zeigt sich, dass Lehrkräfte ihre Interaktion mit den Schülerinnen und Schülern an ihre Erwartungen anpassen und so Erwartungen zu einem kleinen aber substanziellen Anteil selbsterfüllende Prophezeiungen werden können (vgl. den Übersichtsartikel von Jussim & Harber, 2005). Obwohl solche individuellen Pygmalioneffekte gut untersucht sind, gibt es kaum Studien zu Erwartungseffekten auf der kollektiven Ebene. Unsere Studie hat zum Ziel, die Forschung zu Erwartungen von Lehrkräften um Stereotypen zu Schulformen zu erweitern. Es wird der Frage nachgegangen, welchen Inhalt und welche Struktur Stereotypen von Lehrkräften zu Schulformen im deutschen gegliederten Schulsystem haben. Dann wird untersucht, wie diese stereotypischen Beurteilungen zu Selbstwirksamkeitserwartungen und der Wahrnehmung von Schwierigkeiten im Klassenzimmer in Beziehung stehen. Querschnittliche Analysen werden mit den Daten von 341 Lehrkräften aus dem Gymnasium und der Hauptschule durchgeführt, die im Rahmen der Panel Study at the Research School „Education and Capabilities“ in North Rhine-Westphalia (PARS) (siehe Bos et al., 2016) befragt wurden. In explorativen Faktorenanalysen wurde zunächst die Struktur der Stereotypen untersucht. Dann wurden Regressionsanalysen zur Bestimmung von Beziehungen zu Selbstwirksamkeitserwartungen und der Wahrnehmung von Schwierigkeiten im Klassenzimmer berechnet. Die Ergebnisse zeigten, dass – nach Kontrolle der durchschnittlichen kognitiven Fähigkeiten und dem kulturellen Kapital der Schülerinnen und Schüler – die Stereotype der Lehrkräfte signifikante Zusammenhänge zu Selbstwirksamkeitserwartungen und der Wahrnehmung von Schwierigkeiten im Klassenzimmer aufwiesen. Nach einer gründlichen Diskussion der Ergebnisse wird ein kurzer Ausblick auf ein Forschungsprogramm zur Untersuchung von Pygmalioneffekten auf der kollektiven Ebene gegeben.

Schlagworte

Lehrkräfte; Stereotype; Selbstwirksamkeit; Pygmalion; Erwartungen kollektiv

1. Introduction

To judge students in various domains is a typical aspect of teachers' professional activities. These diagnoses are fundamental to making professional decisions about appropriate teaching strategies, which sometimes have to rely on somewhat vague information, especially when a teacher starts to work with a class. This is of particular importance as teacher expectancies can influence students' development. Combining the results of eight meta-analyses (including 674 studies), Hattie (2009) showed a medium average effect of individual teacher expectancies on the actual intellectual development of students ($d = .43$). Nevertheless, effect sizes differ very much between studies: Individual expectancy effects are particularly high when teachers are given misleading information about students' capabilities prior to the substantial direct experience of students (Raudenbush, 1984). Rosenthal and Jacobson (1968) were the first to report results on this so-called Pygmalion effect in the educational field. One important factor in the explanation of the individual Pygmalion effect is that teachers tend to adapt their interaction styles to their expectations (as summarized in a review by Jussim & Harber, 2005), which results in fewer learning opportunities for students with less positive prognoses.

1.1 Collective teacher expectations

There seems to be only little research on teachers' expectations of what their whole class or school is or is not going to achieve. This is especially surprising since Brophy claimed already 25 years ago that "differential teacher treatment of intact groups and classes may well be a much more widespread and powerful mediator of self-fulfilling prophecy effects on student achievement than differential teacher treatment of individual students within the same group or class" (Brophy, 1985, p. 309). A major amount of teaching time consists of interactions between the teacher and the class or other groups as a whole. Since human information processing capacities are limited, teachers are not able to observe all students for their individual characteristics all of the time. It is likely, therefore, that teachers will apply schemes at the group level because this kind of cognitive processing demands less cognitive capacity. Accordingly, teacher expectations regarding entire classes or even schools should have a strong impact on teachers' behaviour in the classroom and could lead to Pygmalion effects based on expectancies at group level. The few existing studies on such collective expectancy effects were mainly conducted by Rubie-Davies and colleagues. Their results indicate that if students believe the teacher to hold low class-level expectations, they may react by viewing themselves

as less academically able (Rubie-Davies, 2006). They also reported that teachers with different class-level expectations showed different interaction styles in the classroom (Rubie-Davies, 2007) and that teachers with high self-reported class-level expectations rated their students' attitude to work and social relationships more favourably (Rubie-Davies, 2010). These studies indicate that constructs at the collective level are promising targets for extended research on Pygmalion effects. Nevertheless, one crucial point is missing in the work of Rubie-Davies and her colleagues: They tried to identify the collective expectations of teachers by aggregating expectations regarding all individual students to a class mean. Accordingly, these studies did not apply direct measures to their investigation of teachers' judgements in the collective dimension. With "direct measurement" of expectancies in the collective dimension we mean that teachers are asked directly how they judge their class or school as a whole. From a social psychological point of view, individual profiles for each student and collective profiles of classes and the school can be assumed to be psychologically distinct dimensions. Since individual and collective expectancies cannot be determined perfectly by each other, they should be assessed as discrete constructs (for an overview of related aspects in social identity theory see Postmes & Branscombe, 2010). We intend to investigate expectancies at the collective level along the lines of this social psychological framework.

1.2 Stereotypes and between-school tracking

Social cognition research addresses the question of how psychological processes influence social interactions (for an overview see Hamilton, 2005). Stereotypes are an important aspect of social cognition. They are defined as attitudes and accordingly have a cognitive, affective and a behavioural dimension (Fiske, 1998). Since stereotypes are attitudes towards groups, they affect especially the collective level of judgements but can also influence those about individuals (Hinnant, O'Brien, & Ghazarian, 2009; McKown & Weinstein, 2002). Stereotypes imply at least one social and one achievement-related dimension and are related to the perceived status of and competition with the outgroup judged (Fiske, Cuddy, Glick, & Xu, 2002). Although stereotypes are – to some degree – defined by the fact that they are socially shared, they also vary to some extent between individuals belonging to the same group. As a consequence, we want to investigate how teachers' stereotypic beliefs about their student groups are related to expectancies regarding obstacles to their teaching as well as self-efficacy in their teaching. Judgements are often biased by stereotypic views. This is especially likely to occur in situations where no comprehensive information is available. Further, people tend to rely on stereotypes more often when they are under a higher cognitive load because the application of stereotypes is an economical means of processing information and regulating action (Fiske, 1998). In stressful situations it is even possible that judgements and behaviour are influenced by stereotypes that are not shared by the person passing the judgement (Dion, 2003; Fazio, 1990). In school settings, a lack of com-

prehensive information can be assumed when teachers start teaching a new class. Similarly, a rather high cognitive load will frequently be present during teaching activities. It is therefore likely that stereotypes about student groups play a role in the formation of teacher expectancies regarding their whole class and school.

An important feature of German schools that conveys rather strong stereotypic content is the school track. The German school system applies between-school tracking. After four or six years at elementary school, students are assigned to different secondary school types according to their prior achievement. Secondary school systems differ between the German federal states. North Rhine-Westphalia, for example, still employs the traditional school tracks *Hauptschule* (lowest), *Realschule*, and *Gymnasium* (highest), extended by *Gesamtschulen* (comprehensive schools). The practice of explicit between-school tracking – placing students in different school types according to their achievement level – is justified by the idea that reducing the heterogeneity in students' capacity makes it easier for teachers to create learning environments that are ideal for the advancement of all of their students (Baumert, Stanat, & Watermann, 2006; Maaz, Trautwein, Lüdtke, & Baumert, 2009). Using similar reasoning, various forms of tracking are applied intentionally in many countries besides Germany. Explicit between-school tracking is common in many European countries such as Belgium, the Netherlands and Switzerland. Between-school tracking can also occur implicitly, as is the case in US-American high schools. Formally, these schools are equivalent regarding intended curriculum and accomplishable degrees. In actual fact, however, high schools differ very strongly depending on the neighbourhood they are set in, which in turn can impact students' ability development, the distribution of degrees attained and the appreciation these degrees have in their particular society. Another form of tracking is within-school tracking which assigns students of the same school to classes or courses with different curricula according to their achievement level. Within-school tracking can similarly occur both implicitly or explicitly. However, there are countries such as Scotland, Sweden, and Norway that have a comprehensive curriculum and no explicit tracking at school or class level (Opdenakker & van Damme, 2006). Although the implementation of achievement criteria is quite accurate in Germany, not all students are tracked appropriately (e.g., Bos et al., 2003). Concurrently, stereotypes about school types are very noticeable and diverse between the school types. They can be reproduced quite consistently by teachers (Knigge, 2009) and students (Knigge & Hannover, 2011) and show substantial relations to motivational foci.

1.3 The present investigation

The classroom situation has an influence on whether classroom interaction processes rely on individual or collective judgement profiles. In more interpersonal situations – for instance, when teachers interact with only two or three children –, it is likely that actions strongly depend on individual-level judgements that teachers hold about students. Nevertheless, a high proportion of student-teacher interactions involve too many individuals to cognitively process the situation mainly on an interpersonal basis. It can be assumed that teacher behaviour will be influenced by his or her judgements in the collective dimension rather than by individual judgements of each individual student, at least in classic instructional situations. Thus, collective judgement profiles are more likely to be relevant for teachers' behaviour. In situations where collective judgements are likely to be influential, stereotypes may come into play and influence classroom processes. This is especially likely in ambiguous situations and when the cognitive load is high: An example of such a situation would seem to be teaching a new class. Accordingly, the influence of stereotypes on the teachers' judgements and teacher-student interactions is likely to be heightened. Following our rationale, we also assumed that the school track functions as one informational source that provides teachers with stereotypes for their judgements in the collective dimension. In the present study, we wanted to investigate the structure of the stereotypes that teachers hold regarding traditional school tracks in Germany. We were, further, interested in how these stereotypes relate to expectancies regarding obstacles to, and efficacy-beliefs about, classroom teaching. It was necessary to control for the influences of the measured cognitive abilities and social backgrounds of the students to ensure that we investigated the influence of stereotypes rather than realistic assumptions.

2. Methods

In this section, we describe the data base and how we dealt with the data.

2.1 Sample

The analyses in our study were carried out on data collected in the *Panel Study at the Research School 'Education and Capabilities' in North Rhine-Westphalia* (PARS) at the Institute for School Development Research (IFS) of the Technological University of Dortmund, Germany. PARS was designed as a longitudinal study in the German federal state of North Rhine-Westphalia as part of the International Research School 'Education and Capabilities' (for further information, see Bos et al., 2016). Secondary schools of all school types in North Rhine-Westphalia participated – among them lower secondary schools (*Hauptschule*), intermediate second-

ary schools (*Realschule*), upper secondary schools (*Gymnasium*), comprehensive schools (*Gesamtschule*), and special needs schools (*Förderschule*). PARS conducted annual written surveys on diverse topics from 2009 to 2012 in all schools with pupils, parents, teachers, head teachers and other teaching staff.

The sample used for our analyses was taken exclusively from the data on lower (*Hauptschule*) and upper (*Gymnasium*) secondary schools which allowed a comparison of two vastly differing German school types (cf. Knigge, 2009; Knigge & Hannover, 2011). The *Hauptschule* prepares for vocational training while the *Gymnasium* has an academic curriculum. Students of the latter come on average from families from a higher socioeconomic background and receive the permission to attend university on the strength of their school-leaving certificate that is usually obtained after 12 or 13 school years. Students at the *Hauptschule* tend to come from socially disadvantaged families and leave school after nine or ten school years. As the public reputation of the *Hauptschule* is rather low, its students often face problems in being admitted to qualified vocational training programmes after graduation. For reasons of economy, teachers of either school type participating in PARS had been asked about stereotypes regarding only their own school type and the *Gymnasium*. Accordingly, we chose the *Gymnasium* and the *Hauptschule* for our analyses as these teachers were the only ones that had been asked for the same stereotypical thoughts, namely what they believed that people stereotypically think about students of the *Hauptschule* and of the *Gymnasium*. This was important for two reasons. First, it ensured that there would not be any differences in the answers due to different contrasts. To illustrate: A teacher might have given a different answer to a question about stereotypes of the *Hauptschule* if she had just indicated her beliefs about stereotypes regarding *special needs schools* immediately before, a different answer that is than if the immediately preceding question had been regarding stereotypes about the *Gymnasium*. As both groups were asked the same items regarding the same two groups we arranged for the contrasts to be identical in both groups. Second, we wanted to use all stereotype items regarding both school types in one comprehensive factor analysis. This would make it possible to identify a common structure in the teachers' stereotypical thoughts that implied not only their absolute beliefs about the stereotypes concerning each school type but also the psychological contrasts the teachers made between the two school types.

We will now focus on the first measurement that took place at the end of 2009 (for students) and in spring 2010 (for teachers) (for further details on the PARS-design, see Bos et al., 2016). These are the figures for the first measurement: There were a total of $N = 50$ schools (of which $n = 7$ were lower (*Hauptschule*) secondary schools; $n = 15$ upper (*Gymnasium*) secondary schools), with $N = 550$ teachers (of which $n = 67$ at lower secondary schools; $n = 274$ at upper secondary schools), and $N = 2133$ pupils (of which $n = 207$ were in 5th grade and $n = 217$ in 9th grade at lower secondary schools; as well as $n = 429$ in 5th grade and $n = 408$ in 9th grade at upper secondary schools). The overall average age of teachers (in number of years) was 43.94 ($SD = 11.23$), with the average age of lower secondary schools

teachers being slightly higher ($M = 49.45$ years, $SD = 8.74$). The average age of upper secondary school teachers was closer to the overall average teacher age ($M = 43.58$ years, $SD = 11.55$). At lower secondary schools, 70.1 % of teachers were women, at upper secondary schools 60.3 %. The students' sociodemographic background also showed differences between school tracks: A majority of students at the *Hauptschulen* were male (53.2 % vs. 48.6 % at the *Gymnasien*) and came from families with an immigrant background (35.6 % vs. 22.1 % at the *Gymnasien*).

2.2 Data preparation

In a first step, we ran an exploratory principal axis factor analysis with the items measuring teachers' stereotypic beliefs regarding the *Hauptschule* (lowest track) and the *Gymnasium* (highest track). Bipolar semantic differential items were used to ask the teachers *what they believed people think students from the Hauptschule/Gymnasium are like*. Note that the teachers were asked what *people* think instead of asking for the teachers' personal beliefs; this was done to avoid biases due to the tendency to fulfil social expectancies. The resulting scales will be described in the section on *Instruments*. A comprehensive description of the factor analysis will be given in the section on *Results*. In a next step, we calculated factor scores by computing the mean of all items representing a factor and then z-standardized these scale scores. Items in reverse scaling were recoded so that a high score on an item would always add to a high level on the corresponding factor.

An analysis of how teachers develop school-track-specific attitudes would be incomplete without controlling for certain student-based variables, such as the cognitive abilities and cultural capital of those students with whom teachers interact on a daily basis in their own school track. However, the PARS dataset does not provide a variable that would have allowed us to match teachers with the students they teach at class level. A compromise was to aggregate scores in cognitive abilities (measured with the test of cognitive abilities *KFT 4-12 + R*; Heller & Perleth, 2000) and the number of books at home as an indicator of students' cultural capital (Bourdieu, 1983) for both 5th and 9th graders. Cultural capital as indicated by the number of books at home is an important mediator for the effects of families' social status on students' performance (cf. Bos, Schwippert, & Stubbe, 2007).

To obtain indicators in order to control for the relative student composition of schools in relation to other schools of the same type, we calculated dummy variables for mean cognitive abilities and mean cultural capital. This was done separately for each school type for students in Grades 5 and 9. We generated dummies indicating whether the schools' 5th or 9th graders' mean cognitive abilities lay in the upper or lower third as compared with their respective reference group (5th graders/9th graders of *Hauptschule/Gymnasium* in the sample). We also calculated two dummy variables indicating whether the schools' 5th or 9th graders' mean cultural capital lay in the upper third or lower third as compared to their reference group. This was considered a more appropriate approach than using the schools'

students' mean scores as predictors since the dataset included only 22 schools (15 *Gymnasien*, 7 *Hauptschulen*), as described in Section 2.1. As regression analyses needed to be conducted separately for each school type, the number of school units was too low to apply the school aggregated mean student measures of cognitive abilities and books at home directly. If the number of units is low (here 15 and 7) it is usual to transform metric information into ordinal information (Bortz & Lienert, 1998). If the number of units is low, biases can occur due to randomly extreme distances between measurement points within the sample if metric scale properties are assumed. Such biases can be avoided through the transformation into an ordinal scale format. Accordingly, the transformation into dummy variables that represent cognitive abilities and cultural capital above and below average was conducted for the sake of the estimation of robust coefficients.

2.3 Instruments

Three scales with respect to teacher expectancies were used as dependent variables: (a) perceived limitation during class, (b) collective teacher self-efficacy, and (c) individual teacher self-efficacy. Scale (a) was originally derived by Helmke and Jäger (2002) to measure limitations that teachers perceive while teaching mathematics classes. For the wider purposes of PARS, the scale was modified to collect information about perceived limitations in class independent of the subject taught. The scale asks for perceived limitations relating to uninterested pupils, students with low socioeconomic status, students with special needs, uninterested parents, students with poor German language skills, etc. On scale (b), teachers rate their collective teacher self-efficacy – as adapted from Schwarzer and Jerusalem (1999) – by rating their expectancies regarding the effectiveness of the teaching staff as a whole (an example statement was: “I believe that the strong innovation potential of my teaching colleagues will help us to cope with adverse circumstances”). Finally, scale (c) focused on individual teacher self-efficacy (Schwarzer & Jerusalem, 1999) as experienced for example in one's own classroom management and contact with parents. Items used were for instance “I know that I can maintain a positive relationship with parents even when tensions arise” (see Table 1).

Table 1: Reliability analyses of all relevant continuous variable

Scale	No. of items	Item-total-correlations	Cronbach's Alpha
Collective teacher self-efficacy	12	.46–.77	.91
Individual teacher self-efficacy	8	.29–.57	.73
Perceived limitation in class	8	.40–.65	.82
Stereotype GY: Cognitive potential	5	.56–.82	.88
Stereotype GY: Discipline	4	.66–.71	.84
Stereotype GY: Personality	5	.43–.54	.71
Stereotype GY: Respectful behaviour	5	.58–.72	.86
Stereotype HS: Behaving well in class	4	.56–.70	.80
Stereotype HS: Cognitive potential	7	.61–.81	.91
Stereotype HS: Humorous personality	2	.37–.37	.54
Stereotype HS: Self-confident	3	.21–.25	.39

2.4 Treatment of missing data

The dataset included missing values on many variables, something that happens not infrequently in the social sciences. In order to avoid problems caused by incomplete data or inadequate methods of dealing with missing data (e.g. loss of statistical power), we adopted the current state-of-the-art approach by applying multiple imputation (Luedtke, Robitzsch, Trautwein, & Koeller, 2007). The multiple imputation technique makes up for the missing values by computing several datasets with various plausible values. Analyses are run separately for each dataset; in a last step, results from all imputed datasets are pooled. For the self-efficacy and limitations-in-the-classroom measures the maximum missing data rate was 25 %. Missing rates for the stereotype measures had a maximum of 25 % at the *Hauptschule* and a maximum of 35 % at the *Gymnasium*. Using the R-based package MICE (Multivariate Imputation by Chained Equations; van Buuren & Groothuis-Oudshoorn, 2011), we imputed ten complete datasets. As indicated by the results of a simulation study conducted by Graham, Olchowski, and Gilreath (2007), ten imputed datasets yield robust results for the missing data rates at hand. Mplus provides a function which analyses several analogously structured datasets simultaneously and pools the results of these datasets in a single step. All models were run using this function, so that all outcomes from regression analyses are based on the ten imputed datasets and their pooled results. Descriptions and intercorrelations are, however, based on the original dataset.

2.5 Regression analyses

We applied linear regression analyses to gain further insight into the data structure. Analyses were run for three criterion variables: collective teacher self-efficacy, individual teacher self-efficacy, and the perceived limitations during class due to student characteristics. All models were run separately by grade (5th grade/9th grade) and by subsample (*Hauptschule/Gymnasium*). Our investigation aimed at the identification of the influence of stereotypes concerning school tracks on expectancies at the collective level. To avoid biases due to specific environments, we opted for separate analyses for each school type. This enabled us to determine whether differences in expectancies arise within fairly similar contexts in covariation with the relative strength of stereotypic beliefs. This strategy made it also possible for us to accept different structures of the stereotypic beliefs depending on the type of school. In order to control for the schools' cognitive level, we included two dummy variables indicating whether a school's 5th graders' (or 9th graders', depending on the subsample analysed) average cognitive abilities lay in the upper third or lower third as compared with their reference group (5th graders/9th graders of the *Hauptschule/Gymnasium* sample). Similarly, we controlled for the schools' cultural capital by including two dummy variables indicating whether the school's 5th graders' (or 9th graders', again depending on the subsample) average cultural capital was located in the upper third or lower third as compared with their reference group (5th graders/9th graders of the *Hauptschule/Gymnasium* sample). In addition to these four dummy variables, all analyses included the four school-type-specific stereotype scores as predictors. These stereotype dimensions of the *Gymnasium* were about students' cognitive potential, discipline, personality and respectful behaviour, and of the *Hauptschule* about cognitive potential, good behaviour in class, humorous personality and self-confidence. The application and results of the factor analysis performed to define these stereotypic dimensions will be reported in detail in the following section. All regression analyses were carried out via Mplus 6 (Muthén & Muthén, 1998–2010). It should be noted that the tables report non-standardized results. For simplicity's sake, all continuous variables were z-standardized.

3. Results

3.1 Exploratory factor analysis

Since the original instrument developed by Knigge (2009) was meant to assess teachers' track-specific stereotypes about *Hauptschule* only, the established factor structure could not be adopted in our analyses of teachers' stereotypes regarding *Hauptschule* and *Gymnasium*. In order to determine the factor structure applicable to teachers' stereotypes, we conducted an exploratory factor analysis. We opted

Table 2: Pattern matrix for all ten factors generated by PAF

Item	Factor									
	1	2	3	4	5	6	7	8	9	10
HS: clever – stupid	-.76	-.04	-.02	-.22	.17	.03	-.07	-.00	.01	.04
HS: intelligent – not intelligent	-.69	-.17	.00	.01	.32	.09	-.04	-.05	-.06	.10
HS: foolish – wise	.69	.02	.08	.15	-.16	.00	.18	.07	.05	-.02
GY: foolish – wise	-.69	-.11	.03	.22	-.06	-.08	.12	.25	.05	-.15
HS: capable – incapable	-.58	-.04	-.04	-.04	.21	.04	-.04	-.11	-.13	-.09
GY: clever – stupid	.57	.18	-.09	-.09	.06	.09	-.06	-.24	-.09	.15
HS: dumb – smart	.56	-.08	-.02	.00	-.27	.04	.14	.04	.09	.01
HS: incompetent – competent	.55	.01	.06	.12	-.09	.13	.14	.18	-.01	.02
GY: dumb – smart	-.50	-.38	-.06	.27	.04	.01	.05	.05	.02	-.17
GY: intelligent – not intelligent	.49	.18	.03	-.27	.17	-.03	.19	-.01	-.07	.20
HS: do not work hard for school – work hard for school	.45	-.18	.06	-.02	-.33	.15	.04	-.14	-.17	.12
GY: do not work hard for school – work hard for school	-.37	-.16	.13	.09	-.04	.04	.04	.21	.10	-.21
HS: do not pay attention during class – pay high attention during class	.30	-.29	.16	.03	-.27	.20	.08	-.05	-.07	.28
GY: listen to their teachers – do not care what their teachers say	.01	.67	.15	-.13	-.12	.00	.08	-.08	-.07	.13
GY: friendly – mean	.21	.57	-.10	.08	-.01	.02	.04	-.10	.04	.17
GY: well-behaved – impudent	.04	.56	.05	-.02	-.12	.18	.10	-.12	-.01	.14
GY: indecent – decent	-.16	-.44	.03	-.17	-.07	-.08	-.06	.18	-.06	-.28
GY: do not pay attention during class – pay high attention during class	-.16	-.43	-.14	.19	.00	-.07	.04	.17	.08	-.16
GY: capable – incapable	.26	.34	-.04	-.18	-.07	.16	-.03	-.01	-.13	.20
GY: cool – uncool	.14	-.09	-.74	.16	.06	.10	-.02	.03	-.11	.00
GY: boring – cheerful	.12	-.07	.69	.04	.04	-.17	-.18	.24	.23	.07
GY: a bore – jokers	.10	.04	.57	.13	.10	.05	-.04	.00	-.08	.09
GY: brave – anxious	.06	.02	-.53	-.00	.14	-.13	-.01	.09	-.22	.21
GY: true to themselves – bootlickers	-.01	-.01	-.32	-.30	-.04	-.16	-.08	-.13	.02	.03
HS: comical – dull	.01	-.26	.29	-.10	-.22	-.20	.20	-.14	.04	-.07
GY: incompetent – competent	.01	-.03	-.01	.55	.02	-.05	-.04	-.11	-.04	-.01

Table 2 continues

Table 2 continued

Item	Factor									
	1	2	3	4	5	6	7	8	9	10
HS: courteous – naughty	-0.14	-0.11	-0.05	.10	.66	-0.06	.03	-0.05	.19	.02
HS: attentive in class – easily distracted	-0.12	-0.15	.09	-0.08	.59	.16	-0.04	.11	-0.08	.01
HS: talk during class – quiet during class	.04	-0.11	.13	-0.01	-.56	.12	.02	-0.13	-0.18	.22
HS: well-behaved – impudent	-.28	-0.06	-0.09	.03	.54	-0.10	-0.04	-0.08	.03	-0.13
HS: lazy – diligent	.17	-0.13	-0.06	-0.02	-.44	.01	.15	.17	.15	.40
HS: listen to their teachers – do not care what their teachers say	-.37	.07	-0.15	-0.02	.38	-0.01	-0.14	-0.04	.12	-0.17
HS: reserved – cheeky	-0.10	.05	-0.05	-0.06	.33	-.28	-0.17	.04	.08	-.32
HS: friendly – mean	-.30	-0.02	-0.06	-.24	.32	-.15	-0.13	-0.03	.20	-0.08
HS: true to themselves – bootlickers	-0.08	.16	-0.14	-0.03	.02	.72	.01	.01	.02	-0.07
HS: brave – anxious	-0.00	-0.14	.14	-0.03	.10	.44	.02	-.24	.22	.01
HS: cool – uncool	.11	-0.07	.05	.04	-.19	.29	-0.06	.06	.12	.05
GY: comical – dull	-0.04	.04	-.12	-0.07	.17	-0.03	.93	-.03	-.03	-0.07
GY: reserved – cheeky	-0.05	.12	-.11	.19	-0.08	.04	.10	-.63	.18	.04
GY: talk during class – quiet during class	.01	-.22	-.27	-.18	.03	-.11	-0.09	.35	.26	-.25
HS: a bore – jokers	-0.06	-0.03	-.11	-.13	-.10	-0.00	.03	.21	-.58	-.09
HS: boring – cheerful	-0.03	.01	-0.06	.20	-0.01	-.15	.03	-0.06	-.58	.01
GY: attentive in class – easily distracted	-0.10	.14	-0.04	-0.07	-.12	.00	-.12	-0.10	-0.04	.78
GY: lazy – diligent	-0.12	-0.05	-0.03	.07	-.14	.03	-0.03	.17	.02	-.71
GY: courteous – naughty	-.26	.30	.08	.04	-0.09	.05	.08	-0.04	.03	.65
GY: thorough – sloppy	.22	.24	-0.05	.06	-0.04	-0.07	.02	.06	.08	.55
HS: thorough – sloppy	-0.09	.08	.02	.01	.44	-.17	-0.08	-0.13	.01	-.47
HS: indecent – decent	.21	.11	.04	.29	-0.07	.11	.27	.10	.05	.37
Eigenwert	5.03	2.45	2.19	1.16	2.88	1.35	1.33	1.22	1.25	3.20
% variance	20.9	10.2	09.1	04.8	12.0	5.6	5.5	5.1	5.2	13.3

Note. All items rating students from a Hauptschule are marked in light grey. All other items rate students from a Gymnasium. Bold-printed values indicate an item's highest factor loading.

for principal axis factoring (PAF) and direct oblimin rotation via SPSS.20, analysing items rating students from both school tracks simultaneously.

PAF generated ten factors with an Eigenvalue greater than 1. Table 2 reports item loadings on all ten factors. Items were assigned to the factor they loaded highest on. Since some items did not load very highly on any factor, we applied the Fürntratt criterion which demands that an item should only be assigned to a factor if its squared loading on this factor accounts for at least 50 % of the item's communality (Bühner, 2011). We excluded items from our final factor solution if they did not even meet a more liberal 40 % version of the Fürntratt criterion.

Two of the ten factors with an Eigenvalue greater than 1 were ignored since they each had only one item assigned to them. Six of the seven remaining factors distinguished quite neatly between tracks so that three factors were exclusively based on items rating *Hauptschule* students and three factors were almost exclusively based on items rating *Gymnasium* students (with the exception of one item). The seventh factor (*cognitive potential*) contained items rating students from both tracks. In order to obtain a more distinct factor solution, it was decided to artificially divide this mixed factor by track, rendering two instead of one factor. The single *Hauptschule* item in one of the *Gymnasium* factors (*discipline*) was also removed.

Table 3: Final factor solution including factor titles, number of items and exemplary items

Factor	No. of items	Examples
Hauptschule		<i>People think students from a Hauptschule to be ...</i>
Cognitive potential	7	<i>... intelligent – not intelligent</i> <i>... stupid – clever</i>
Appropriate behavior in class	4	<i>... attentive in class – easily distracted</i> <i>... well-behaved – cheeky</i>
Self-confident	3	<i>... true to themselves – bootlickers</i> <i>... brave – anxious</i>
Humorous personality	2	<i>... a bore – jokers</i> <i>... boring – cheerful</i>
Gymnasium		<i>People think students from a Gymnasium to be ...</i>
Cognitive potential	5	<i>... intelligent – not intelligent</i> <i>... stupid – clever</i>
Respect towards teachers	5	<i>... listen to their teachers – do not pay attention to their teachers</i> <i>... friendly – mean</i>
Personality	5	<i>... a bore – jokers</i> <i>... true to themselves – bootlickers</i>
Discipline	4	<i>... lazy – diligent</i> <i>... thorough – sloppy</i>

This procedure led to a clean-cut factor solution with four factors for each track. Table 3 summarizes this final factor solution, namely for the *Hauptschule* cognitive potential, appropriate behaviour in class, self-confidence and humorous personality, and for the *Gymnasium* cognitive potential, respect towards teachers, personality and discipline. Note that the example items are reported in their original, not yet recoded version. They were, of course, recoded later so that a high score on any factor would indicate a positive manifestation of this variable (thus, a high score on the factor *cognitive potential – Hauptschule* would in fact indicate that a teacher believed that other people think of *Hauptschule* students as intelligent and clever).

3.2 Descriptives

Descriptives were calculated for teachers from both school tracks together. Means, standard deviations and the number of cases with valid data for all individual variables are presented in Table 4. Intercorrelations are reported separately for *Gymnasium* teachers and *Hauptschule* teachers in Table 5.

The self-efficacy means indicate that teachers, on the whole, feel rather confident and self-efficient both as individual teachers and as staff members (individual self-efficacy: $M = 2.99$, $SD = 0.39$; collective self-efficacy: $M = 2.97$, $SD = 0.45$; scales ranging from 1 to 4), with the two scales being positively correlated in both teacher groups. Teachers did not, on the whole, perceive themselves as being limited during class by student characteristics ($M = 2.13$, $SD = 0.60$; scale ranging from 1 to 4).

At an average descriptive level, students from the *Gymnasium* were rated rather positively for their cognitive abilities, discipline and respectful behaviour (stereotype GY [= *Gymnasium*] cognitive potential: $M = 1.82$, $SD = 0.83$; stereotype GY discipline: $M = 1.37$, $SD = 0.89$; stereotype GY respectful behavior: $M = 1.29$, $SD = 0.86$; scales ranging from -3 to +3) but neither positively nor negatively regarding their personality. Likewise, mean ratings of *Hauptschule* students' personality traits (humorous, self-confident) were neither particularly positive nor negative. However, when it came to cognitive potential and appropriate behaviour in class, students from the *Hauptschule* were, on average, rated negatively (stereotype HS [= *Hauptschule*] cognitive potential: $M = -1.52$, $SD = 0.99$; stereotype HS behaving well in class: $M = -1.63$, $SD = 1.00$; stereotype GY respectful behavior: $M = 1.29$, $SD = 0.86$; scales ranging from -3 to +3).

Individual and collective teacher self-efficacy share small positive correlations with three stereotypical dimensions in the *Gymnasium* subsample; in the *Hauptschule* subsample, collective teacher self-efficacy shows a moderate negative correlation with rating *Hauptschule* students as stereotypically self-confident. The degree to which a teacher feels limited in class due to student characteristics is unrelated to self-efficacy variables or stereotypical ratings in the *Hauptschule* subsample; in the *Gymnasium* subsample, perceived limitation in class correlates

positively with rating *Gymnasium* students as cognitively able and behaving respectfully.

Table 4: Descriptive statistics

Scale	<i>N</i>	<i>M</i>	<i>SD</i>
Collective teacher self-efficacy	265	2.97	0.45
Individual teacher self-efficacy	253	2.99	0.39
Perceived limitation in class	327	2.13	0.60
Stereotype GY: Cognitive potential	230	1.82	0.83
Stereotype GY: Discipline	231	1.37	0.89
Stereotype GY: Personality	228	0.18	0.76
Stereotype GY: Respectful behaviour	228	1.29	0.86
Stereotype HS: Behaving well in class	235	-1.63	1.00
Stereotype HS: Cognitive potential	233	-1.52	0.99
Stereotype HS: Humorous personality	229	0.11	0.93
Stereotype HS: Self-confident	228	0.36	0.82

Note. Stereotype scales range from -3 to +3, perceived limitations and self-efficacy beliefs range from 1 to 4.

3.3 Regression analyses

Regression analyses were conducted separately for both school types and grades for all three criterion variables individual teacher efficacy, collective teacher efficacy and perceived limitations in class. To ensure that results were not biased by multicollinearity between stereotype dimensions, models were computed with one dimension solo for each stereotype dimension. Multicollinearity issues with the dummy variables reflecting the schools' average cognitive abilities and cultural capital were avoided by computing models with and without these variables. Finally, comprehensive models including all variables were computed. The results for all these models are summarized in six tables (3 criterion variables x 2 school types) named 6a to 6f.

Table 5: Intercorrelations

Scale	Collective teacher self-efficacy	Individual teacher self-efficacy	Perceived limitation in class	GY: Cognitive potential	GY: Discipline	GY: Personality	GY: Respectful behaviour	HS: Behaving well	HS: Cognitive potential	HS: Humorous personality	HS: Self-confident
Collective teacher self-efficacy	.39	-.08	-.08	.26	.20	.11	.18	-.03	-.01	.09	.18
Individual teacher self-efficacy	.57	.00	.00	.16	.16	.12	.23	-.04	.09	.02	.03
Perceived limitation in class	-.21	-.23		.28	.06	-.05	.15	-.25	-.16	-.08	.00
Stereotype GY: Cognitive potential	.10	.23	.16		.69	.09	.69	-.45	-.53	-.02	.18
Stereotype GY: Discipline	.04	.23	.02	.74		.02	.69	-.38	-.36	.07	.18
Stereotype GY: Personality	-.22	-.16	.10	.03	-.12		.16	.19	.17	-.16	-.07
Stereotype GY: Respectful behaviour	.09	.20	.24	.77	.72	-.14		-.36	-.37	-.05	.19
Stereotype HS: Behaving well in class	.05	-.03	.01	-.41	-.49	.08	-.42		.68	.02	-.22
Stereotype HS: Cognitive potential	.08	.06	.02	-.44	-.45	.12	-.44	.69		.10	-.12
Stereotype HS: Humorous personality	.14	.09	.01	-.06	-.06	-.34	.13	-.14	-.08		.36
Stereotype HS: Self-confident	-.42	-.24	-.14	-.02	.05	.05	.04	.09	.12	.05	

Note. Statistically significant correlations ($p < .05$) are printed in boldface. Values above the diagonal report intercorrelations in the Gymnasium teacher subsample; values below the diagonal report intercorrelations in the Hauptschule teacher subsample.

Table 6a: Regression analyses: Individual teacher self-efficacy as criterion variable; subsample Gymnasium

Predictor	Gymnasium – 5th grade					Gymnasium – 9th grade														
	1	2	3	4	5	1	2	3	4	5										
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE										
Dummy: CogAbi lower third	-.46*	.21	-.44*	.21	-.44*	.21	-.41*	.21	-.00	.22	-.03	.22	.02	.22	.01	.22	-.04	.22		
Dummy: CogAbi upper third	-.30	.20	-.29	.20	-.33	.22	-.25	.21	-.27	.20	.20	.20	-.32	.19	.26	.20	.21	.20		
Dummy: SociStat lower third	-.10	.22	-.12	.22	-.13	.23	-.15	.22	-.12	.21	.19	.15	.19	.20	.19	.20	.19	.15	.19	
Dummy: SociStat upper third	-.26	.21	-.26	.21	-.28	.22	-.27	.21	-.25	.20	.11	.21	.06	.21	.11	.21	.11	.21	.06	.21
Stereotype 1: CogPot	.34**	.09					.19	.14	.17*	.09									.06	.14
Stereotype 2: RespecBehav			.37**	.10			.20	.15			.21**	.08							.21	.13
Stereotype 3: Personality					.12	.11			.15	.10			.09	.09					.03	.09
Stereotype 4: Discipline					.34**	.11	.05	.17							.13	.09	-.07			.14
R² – only Stereotype	.12*	.06	.12	.06	.02	.03	.10	.06	.03	.03	.06	.04	.01	.01	.02	.03				
R² – all predictors	.18*	.07	.18*	.07	.08	.05	.15*	.07	.22**	.07	.06	.04	.07	.04	.04	.03	.05	.04	.08	.04

Note. CogAbi = Cognitive abilities; SociStat = Social status, CogPot = Cognitive potential; RespecBehav = Respectful Behavior. Analyses are restricted to the Gymnasium subsample and were executed separately for 5th grade and 9th grade teachers, respectively. Individual teacher self-efficacy is the criterion variable; predictor variables are listed in the left-hand column. Dummy variables are coded in a way that 1 = school belongs to the named group (e.g., cognitive abilities – lower third) and 0 = school does not belong to the named group. Analyses were conducted with ten imputed datasets. All variables except the dummies are z-standardized. ** regression weight is statistically significant at $p < .01$. * regression weight is statistically significant at $p < .05$.

Table 6b: Regression analyses: Individual teacher self-efficacy as criterion variable; subsample Hauptschule

Predictor	Hauptschule – 5th grade					Hauptschule – 9th grade																		
	1	2	3	4	5	1	2	3	4	5														
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE														
Dummy: CogAbi lower third	.60	.45	.46	.46	.33	.45	.53	.45	.24	.45	-.20	.57	-.19	-.57	-.26	.57	-.21	.59	-.26	.59				
Dummy: CogAbi upper third	-.18	.42	-.15	.41	.07	.40	-.25	.42	-.05	.41	-.27	.45	-.26	.44	-.29	.44	-.27	.45	-.32	.47				
Dummy: SociStat lower third	-.45	.46	-.45	.46	-.03	.48	-.46	.45	-.02	.47	-.15	.35	-.15	.35	-.16	.34	-.15	.34	-.17	.35				
Dummy: SociStat upper third	-.88	.46	-.90	.47	-.43	.50	-.87	.45	-.41	.49	-.28	.46	-.28	.46	-.28	.46	-.29	.48	-.32	.50				
Stereotype 1: CogPot	.04	.16							.11	.20	-.01	.19								-.05	.28			
Stereotype 2: Behaving well in class			-.01	.19					-.16	.24			-.01	.16							-.02	.23		
Stereotype 3: Self-confident																					.13	.20		
Stereotype 4: Humorous																					-.01	.16	-.03	.18
R² – only Stereotype	.00	.02	.01	.03	.19	.13	.04	.07			.00	.02	.00	.01	.01	.03	.00	.00	.00	.00				
R² – all predictors	.16	.12	.16	.12	.25	.14	.18	.13	.30*	.14	.02	.04	.02	.04	.03	.05	.02	.04	.03	.06				

Note. CogAbi = Cognitive abilities; SociStat = Social status, CogPot = Cognitive potential. Analyses are restricted to the Hauptschule subsample and were executed separately for 5th grade and 9th grade teachers, respectively. Individual teacher self-efficacy is the criterion variable; predictor variables are listed in the left-hand column. Dummy variables are coded in a way that 1 = school belongs to the named group (e.g., cognitive abilities – lower third) and 0 = school does not belong to the named group. Analyses were conducted with ten imputed datasets. All variables except the dummies are z-standardized. ** regression weight is statistically significant at $p < .01$. * regression weight is statistically significant at $p < .05$.

Table 6c: Regression analyses: Collective teacher self-efficacy as criterion variable; subsample Gymnasium

Predictor	Gymnasium – 5th grade					Gymnasium – 9th grade										
	1	2	3	4	5	1	2	3	4	5						
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE						
Dummy: CogAbi lower third	-.41*	.20	-.40*	.20	-.43*	.21	-.33	.20	-.35	.20	-.06	.21	-.06	.21	-.06	.21
Dummy: CogAbi upper third	.03	.20	.03	.20	-.01	.21	.11	.20	.09	.20	.52**	.18	.51**	.19	.57**	.18
Dummy: SociStat lower third	-.02	.21	-.04	.21	-.04	.22	-.08	.21	-.06	.21	-.11	.18	-.14	.18	-.12	.18
Dummy: SociStat upper third	-.27	.20	-.28	.21	-.30	.22	-.29	.20	-.28	.20	-.10	.20	-.12	.20	-.10	.20
Stereotype 1: CogPot	.34**	.09							.19	.14	.18*	.08				.16
Stereotype 2: RespecBehav			.30**	.10					-.04	.15	.10 ^a	.08				-.11
Stereotype 3: Personality									.01	.10						.06
Stereotype 4: Discipline									-.03	.10						-.06
									.43**	.10	.29	.17				.17*
R² – only Stereotype	.13*	.06	.08	.05	.00	.00	.14*	.07			.04	.04	.03	.03	.00	.01
R² – all predictors	.23**	.07	.18*	.07	.10	.06	.24**	.07	.26**	.08	.15**	.06	.13*	.02	.12*	.05

Note. CogAbi = Cognitive abilities; SociStat = Social status, CogPot = Cognitive potential; RespecBehav = Respectful Behavior. Analyses are restricted to the Gymnasium subsample and were executed separately for 5th grade and 9th grade teachers, respectively. Collective teacher self-efficacy is the criterion variable; predictor variables are listed in the left-hand column. Dummy variables are coded in a way that 1 = school belongs to the named group (e.g., cognitive abilities – lower third) and 0 = school does not belong to the named group. Analyses were conducted with ten imputed datasets. All variables except the dummies are z-standardized.

^a Stereotype “Respectful behavior” is not significant when entered into the analysis as the only predictor ($b = 0.15, p = 0.063$); when combined with the dummy variables concerning cultural capital, it reaches statistical significance ($b = 0.16, p = 0.049$). This effect subsides when the dummy variables indicating cognitive abilities are added as predictors ($b = 0.10, p = 0.211$).

** regression weight is statistically significant at $p < .01$. * regression weight is statistically significant at $p < .05$.

Table 6d: Regression analyses: Collective teacher self-efficacy as criterion variable; subsample Hauptschule

Predictor	Hauptschule – 5th grade					Hauptschule – 9th grade										
	1	2	3	4	5	1	2	3	4	5						
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE						
Dummy: CogAbi lower third	1.31*	.60	1.36*	.62	1.08	.62	1.12*	.57	.82	.54	-1.35*	.63	-1.40*	.65	-1.33*	.65
Dummy: CogAbi upper third	-.19	.55	-.06	.56	.17	.55	-.36	.52	-.25	.49	-.52	.50	-1.12*	.51	-.54	.49
Dummy: SociStat lower third	-.15	.60	-.17	.62	.26	.67	-.20	.56	.27	.56	-.39	.38	-1.31*	.64	-.43	.38
Dummy: SociStat upper third	-.65	.61	-.71	.63	-.24	.69	-.64	.57	-.16	.59	-1.05*	.51	-1.12*	.51	-1.23*	.53
Stereotype 1: CogPot	.23	.21							.46	.24	.15	.21				
Stereotype 2: Behaving well in class			.01	.25					-.42	.29		.05	.17			
Stereotype 3: Self-confident					-.32 ^a	.23			-.39*	.19				-.14	.20	
Stereotype 4: Humorous							.36*	.15	.39**	.14				-.12	.18	-.12
R² – only Stereotype	.02	.05	.01	.03	.11	.11	.15	.12			.00	.02	.01	.02	.04	.00
R² – all predictors	.33*	.14	.30*	.14	.35*	.14	.41**	.14	.53**	.13	.15	.11	.14	.11	.15	.11

Note. CogAbi = Cognitive abilities; SociStat = Social status, CogPot = Cognitive potential. Analyses are restricted to the Hauptschule subsample and were executed separately for 5th grade and 9th grade teachers, respectively. Collective teacher self-efficacy is the criterion variable; predictor variables are listed in the left-hand column. Dummy variables are coded in a way that 1 = school belongs to the named group (e.g., cognitive abilities – lower third) and 0 = school does not belong to the named group. Analyses were conducted with ten imputed datasets. All variables except the dummies are z-standardized.

^a Stereotype “Self-confident” is not significant when entered into the analysis as the only predictor ($b = -.41, p = 0.061$); when combined with the dummy variables concerning cultural capital, it reaches statistical significance ($b = -0.48, p = 0.018$). This effect subsides when the dummy variables indicating cognitive abilities are added as predictors ($b = -.32, p = 0.149$).

** regression weight is statistically significant at $p < .01$; * regression weight is statistically significant at $p < .05$.

Table 6c: Regression analyses: Perceived limitations during class as criterion variable; subsample Gymnasium

Predictor	Gymnasium – 5th grade					Gymnasium – 9th grade													
	1	2	3	4	5	1	2	3	4	5									
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE									
Dummy: CogAbi lower third	.18	.26	.16	.26	.15	.26	.03	.25	-.78**	.24	-.80**	.24	-.74**	.24	-.77**	.24	-.79**	.24	
Dummy: CogAbi upper third	.37	.26	.37	.26	.35	.26	.25	.25	-.56**	.21	-.32**	.21	-.49*	.21	-.57**	.21	-.61**	.21	
Dummy: SociStat lower third	.01	.27	-.00	.28	.01	.28	.01	.28	.09	.21	.03	.21	.10	.21	.08	.21	.06	.21	
Dummy: SociStat upper third	-.06	.27	-.06	.27	-.07	.27	-.07	.27	-.12	.23	-.16	.23	-.11	.23	-.12	.23	-.15	.23	
Stereotype 1: CogPot	.14	.12					.40*	.18	.23**a	.10							.18	.16	
Stereotype 2: RespecBehav	.10	.14					.18	.19			.21**b	.10					.15	.14	
Stereotype 3: Personality			-.08	.14			-.14	.13					-.04	.10			-.10	.10	
Stereotype 4: Discipline					-.11	.15	-.62**	.22					.18	.10	-.06	.16			
R² – only	.02	.03	.01	.02	.01	.02	.01	.02	.03	.03	.02	.03	.00	.01	.02	.02			
R² – all predictors	.04	.04	.03	.04	.03	.03	.04	.12	.06	.13*	.05	.12*	.05	.09	.05	.11*	.05	.14*	.06

Note. CogAbi = Cognitive abilities; SociStat = Social status, CogPot = Cognitive potential; RespecBehav = Respectful Behavior. Analyses are restricted to the Gymnasium subsample and were executed separately for 5th grade and 9th grade teachers, respectively. Perceived limitations during class is the criterion variable; predictor variables are listed in the left-hand column. Dummy variables are coded in a way that 1 = school belongs to the named group (e.g. cognitive abilities – lower third) and 0 = school does not belong to the named group. Analyses were conducted with ten imputed datasets. All variables except the dummies are z-standardized.

^a Stereotype “Cognitive potential” is not significant when entered into the analysis as the only predictor ($b = 0.20, p = 0.060$); when combined with the dummy variables concerning cognitive abilities, it reaches statistical significance ($b = 0.23, p = 0.018$). It remains significant when the dummy variables indicating cultural capital are added as predictors ($b = 0.23, p = 0.018$). ^b Stereotype “Respectful behavior” is not significant when entered into the analysis as the only predictor ($b = 0.16, p = 0.082$) or when combined with the dummy variables concerning cultural capital ($b = 0.16, p = 0.094$). It reaches statistical significance when the dummy variables indicating cognitive abilities are added as predictors ($b = 0.21, p = 0.021$).

** regression weight is statistically significant at $p < .01$. * regression weight is statistically significant at $p < .05$.

Table 6f: Regression analyses: Perceived limitations during class as criterion variable; subsample Hauptschule

Predictor	Hauptschule – 5th grade										Hauptschule – 9th grade									
	1		2		3		4		5		1		2		3		4		5	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Dummy: CogAbi lower third	-.51	.43	-.52	.44	-.50	.46	-.50	.44	-.51	.47	.14	.58	.01	.59	.28	.54	.20	.60	.45	.53
Dummy: CogAbi upper third	-.10	.40	-.10	.40	-.09	.41	-.08	.40	-.10	.43	.13	.46	.23	.46	.34	.42	.29	.46	.26	.42
Dummy: SociStat lower third	-.38	.44	-.38	.44	-.39	.49	-.39	.44	-.39	.49	.22	.35	.30	.36	.32	.32	.26	.35	.27	.31
Dummy: SociStat upper third	.29	.44	.30	.45	.28	.50	.28	.44	.29	.51	.04	.47	.16	.48	.14	.43	.34	.49	.25	.45
Stereotype 1: CogPot	.02	.15							.00	.20	-.27	.19							-.33	.25
Stereotype 2: Behaving well in class			.04	.18					.04	.25			-.02	.16					.39	.21
Stereotype 3: Self-confident					.00	.16			.01	.17					-.49**	.17			-.48**	.18
Stereotype 4: Humorous											-.01	.12	-.01	.12			.19	.17	.28	.16
R² – only Stereotype	.00	.00	.00	.02	.00	.02	.01	.04			.07	.08	.00	.02	.16	.11	.04	.07		
R² – all predictors	.26	.14	.26	.14	.26	.14	.26	.14	.26	.14	.09	.09	.04	.06	.21	.12	.07	.08	.31*	.13

Note. CogAbi = Cognitive abilities; SociStat = Social status, CogPot = Cognitive potential. Analyses are restricted to the Hauptschule subsample and were executed separately for 5th grade and 9th grade teachers, respectively. Perceived limitations during class is the criterion variable; predictor variables are listed in the left-hand column. Dummy variables are coded in a way that 1 = school belongs to the named group (e.g., cognitive abilities – lower third) and 0 = school does not belong to the named group. Analyses were conducted with ten imputed datasets. All variables except the dummies are z-standardized. ** regression weight is statistically significant at $p < .01$. * regression weight is statistically significant at $p < .05$.

Table 6a shows the results on individual teacher self-efficacy for the *Gymnasium*. In Grades 5 and 9, positive effects can be identified for the cognitive potential and the respectful behaviour stereotype dimension. These effects were substantially larger in 5th grade. Additionally, the stereotypic beliefs regarding discipline showed a positive effect in 5th grade while this turned out to be insignificant in 9th grade. However, in the comprehensive model these effects were not significant anymore. The corresponding results regarding individual teacher self-efficacy for the *Hauptschule* are listed in Table 6b. A surprisingly strong effect was found in 5th grade, showing that individual teacher self-efficacy lowers if students of the *Hauptschule* are seen as rather self-confident. This effect was also significant in the comprehensive model. In interpreting these results it needs to be taken into account that the reliability of the self-confidence dimension was rather critical. In 9th grade no effects were statistically significant.

Tables 6c and 6d summarize how collective teacher self-efficacy related to the stereotype dimensions in the *Gymnasium* and the *Hauptschule*, respectively. As for individual teacher self-efficacy, more and stronger effects were found in 5th than in 9th grade. For the 5th grade of the *Gymnasium*, medium sized positive effects were found for the stereotypes cognitive potential, respectful behaviour, and discipline. In 9th grade, the same effects were significant but rather small in effect size. The comprehensive model did not reveal any significant stereotype dimensions, indicating that these individual effects could have a common basis. For the *Hauptschule*, significant results were found in 5th grade only. A positive medium size effect was found for the stereotype dimension *humorous*, indicating that teachers at the *Hauptschule* experience more collective teacher self-efficacy if they see their students as rather active and funny. This effect was also significant in the comprehensive model. However, it should not be forgotten that the reliability of the *humorous* dimension was marginal. On the other hand, the self-confidence stereotype showed a small to medium negative effect if modelled separately. Here it is important to remember that the reliability of the self-confidence dimension was critical. Although this effect reached significance in the comprehensive model, it was not significant in all models.

Finally, the results on the perceived limitations in class are reported in Table 6e for the *Gymnasium* and Table 6f for the *Hauptschule*. For the *Gymnasium*, individual effects were only significant in 9th grade. Surprisingly, small effects manifested themselves in those cases where the perception of limitations was stronger when the stereotypical view of *Gymnasium* students showed a tendency towards higher cognitive ability and respectful behaviour. Nevertheless, these effects did not reach significance across all models computed or in the comprehensive model. As the dummies for cognitive abilities indicate, teachers in the 9th grade of the *Gymnasium* see more limitations in classes if they teach at average schools (indicated by the significant negative effects of the two dummies for schools that are above or below average). In 5th grade, significant effects only occurred in the comprehensive model. As before, there was a positive effect of cognitive potential. Additionally, a negative effect for discipline was significant with a considerably

high regression weight of $B = -.62$. This indicates that once the *Gymnasium* stereotype regarding discipline is one standard deviation higher, the perception of limitations is on average 62 % of a standard deviation lower. At the *Hauptschule*, the dimension of self-confidence reached significance only in 9th grade with a small to medium effect size. Here too it needs to be considered that the self-confidence dimension's reliability was critical. The results indicated, that the more teachers of the *Hauptschule* stereotypically see self-confidence in their students in 9th grade, the less they perceive limitations in class.

4. Discussion

In this section, we will first discuss the findings on the factor structure and descriptive results before we go on to address the main research question of this article – how to interpret the results with regard to a relationship between collective expectancies and school track stereotypes. Third, strengths and limitations of the study at hand will be specified. In a final section, the need for a research programme on collective Pygmalion effects will be established followed by a sketch of the major milestones of such an endeavour.

4.1 The structure of teachers' stereotypes about school tracks

In our study, we included teachers from two of those traditional school tracks that would achieve the strongest stereotypical contrast, *Hauptschule* (lowest secondary track) and *Gymnasium* (highest secondary track). Teachers from both school tracks were asked to rate stereotypic views about both school tracks. A previous study had investigated the structure of teacher stereotypes regarding school tracks for teachers at the *Hauptschule* only (Knigge, 2009). Accordingly, we decided to search for an integrated solution considering the data for *Hauptschule* and *Gymnasium* at the same time. We identified a common stereotype structure with four dimensions each for *Hauptschule* and *Gymnasium* students. This solution can powerfully explain differential answers to most of the items included (as can be seen by the large amount of variance explained). All dimensions describing stereotypes of the *Gymnasium* showed good or at least satisfactory measurement properties. Two of the dimensions describing *Hauptschule* stereotypes showed good to excellent measurement properties. One dimension, humorous personality, was marginal and one, self-confidence, should be considered as critical. It is possible that these rather mixed findings for the factor structure of stereotypes about the *Hauptschule* result partly from the much smaller sample of *Hauptschule* teachers. It is also possible that the items used to assess these dimensions were not coherent enough to capture the common underlying stereotypic attitude. Clearly, a closer look at these dimensions of stereotypes about *Hauptschule* students is needed, which also takes

into account that teacher stereotypes may vary between federal states. This argument is supported by differences in the stereotypic structure found in this study and the results of the study reported earlier for Berlin teachers (Knigge, 2009).

Still, our approach revealed that the stereotypic descriptions of the two school tracks are quite distinct. As the factor loadings (Table 2) show, only stereotypic beliefs about the cognitive potential of students of the *Hauptschule* and the *Gymnasium* seem to be structurally related. Teachers who believe stereotypically in a relatively higher cognitive potential of students at the *Hauptschule* express relatively lower stereotypic beliefs of the cognitive potential of students at the *Gymnasium*. This indicates that there are some teachers who assume greater and some teachers who assume smaller differences in the cognitive potential between students of the two school types. With regard to all other aspects, most teachers appear to hold structurally differentiated stereotypes regarding *Hauptschule* and *Gymnasium*. Regarding the dimensional structure within both groups, our findings are in accordance with the meta-structure of stereotypes identified by Fiske et al. (2002) in that stereotypes always consist of at least one achievement dimension and one social dimension. For both groups we found clear achievement and social dimensions. However, the social aspect was more differentiated with three dimensions per school track. And our results indicate that the stereotype dimensions of the social aspect are not equal across the two groups.

In summary, a feasible solution was found for the investigation of teachers' stereotypic beliefs about the two school tracks of *Hauptschule* and *Gymnasium*. Students of the *Hauptschule* were stereotypically judged negatively with regard to cognitive potential and behaviour in the classroom. *Gymnasium* students, in contrast, received positive stereotypic judgements regarding cognitive potential, discipline and respectful behaviour. The remaining, more personality-based dimensions were judged rather neutrally for both student groups. Thus, it seems reasonable to apply the identified stereotype structure to the investigation of whether teachers' school-track-specific stereotypes about their own track relate to self-efficacy beliefs and the perception of obstacles in the classroom. The results of this enquiry will be discussed in the following section.

4.2 Relationships between teachers' school track stereotypes and expectancies at the class level as the perception of obstacles in the classroom and efficacy-beliefs

The main question of this study is how teachers' stereotypic thoughts about their school track are related to their expectations that they will be successful teachers. This question was derived from the assumption that if there was such a connection it could lead to Pygmalion effects for instance at the class- or school-level. We investigated whether stereotypic beliefs regarding one's own school track are related to teachers' individual and collective self-efficacy beliefs and the perception of obstacles to teaching. If teachers expect their teaching attempts to be unlike-

ly to succeed, it can be assumed in accordance with the rational choice theory that their efforts will be lowered (e.g., Jonsson, 1999; Wigfield & Eccles, 2000). Thus, such beliefs could become self-fulfilling prophecies. In our investigation we were able to identify substantial relationships between stereotypic beliefs and expectancy-related measures. By taking into account the students' average cognitive abilities and average cultural capital at the school-level, we controlled for central aspects of teachers' accurate diagnoses about their students. Accordingly, the final effects within the regression models reflect a relationship between stereotypes and collective expectancies, hardly influenced by real experiences with the classes involved. Due to the structure of the PARS data, we were not able to apply this control at the class-level and cannot, therefore, completely exclude influences of class diagnostics. But if classes within schools are rather heterogeneous, it would lower the test's power if we were only controlling for school-level effects. Accordingly, our strategy can be assessed as fairly conservative and thus significant effects can be considered to be interpretable. This limitation is one aspect that future studies should address by making the linkage between teacher and student data possible at the class level.

We hypothesized that stereotypes would be especially likely to play a role in the formation of expectancies when there is a scarcity of real information. In line with that assumption, we found that effects on individual and collective teacher-expectancies were prominently present in 5th grade, when teachers had little real experience of their students and thus relied more strongly on other available information, such as their personal stereotypes. The identified relationships at the *Gymnasium* fit very well with our assumptions: The more teachers hold stereotypes that their students are cognitively able, well behaved and disciplined, the more they experience individual and collective self-efficacy. This was especially the case in 5th grade, when there was an almost complete absence of information based on experience with one's students. The *Hauptschule* results revealed a rather complex picture. In line with our assumptions, collective self-efficacy in 5th grade became stronger if the students were rated fairly humorous. The reliability, however, of this dimension was marginal so that this result should be handled with care. What was unexpected was that the intensity of the stereotype that students at the *Hauptschule* are self-confident showed negative effects on both the individual and the collective self-efficacy beliefs in 5th grade. Two aspects have to be considered in the interpretation of this result: First, being the only scale used in this study, the *self-confident* scale showed a critically low reliability (Cronbach's Alpha = .39). Second, it might be possible that the semantic differential items carry additional information within this scale that is not covered by its label and that the effect is produced by strong covariation with only one or two of the items. As the low reliability of a scale leads to more noise in the measurement, this does not really explain the effect. It is rather surprising that significant results are found despite these bad measurement conditions. To analyse whether the second aspect might be important, we reviewed the items of the *self-confident* scale again. The three items consist of the opposite adjective pairs *true to themselves* vs. *bootlickers*, *brave* vs. *anxious*, and *cool* vs. *uncool*. There may well be a perception of sto-

lidity hidden within this construct that implies an association of rather low malleability for these students.

Regarding the perception of limitations in class, results were different. Effects turned out to be significant in 9th grade rather than in 5th grade and their directions were partly unexpected. At the *Gymnasium*, the perceived limitations in class related positively to the strength of the stereotypes about *Gymnasium* students having highly respectful behaviour in 9th grade. However, this effect could neither be identified in the model considering only the stereotype without controlling for student composition characteristics nor in the comprehensive model, quite apart from the fact that the effect was small anyway. In 5th grade at the *Gymnasium*, it was only in the comprehensive model that substantial effects, which were positive for cognitive potential but negative for discipline, reached a significant level. Teachers holding the stereotypic belief that *Gymnasium* students are disciplined perceived substantially less limitations in class. What was, however, unexpected was the direction of the *cognitive potential* belief: The stronger it was the more obstacles the teachers saw. This effect could possibly be explained by the presence of very skilled students causing trouble. Another reason could be that teachers at the *Gymnasium* see themselves confronted with a growing heterogeneity of their classes. Finally, this effect might reflect teachers' doubts about possessing the necessary knowledge and competencies to offer all students sufficient learning opportunities. As these effects are all rather elusive they should clearly be interpreted with a certain amount of caution.

At the *Hauptschule*, a negative effect was found only for the stereotypic self-confidence assumption in 9th grade. For an explanation of this, it should be remembered that the reliability of the self-confidence dimension must be considered critical. Nevertheless, as this effect was medium in size and stable throughout all computed models, it should be taken seriously. It can be the case that some items of the self-confidence factor took on an assumed aspect of maturity that was perceived as positive for a well-functioning classroom by the teachers in the higher grades of the *Hauptschule*. But it is not possible to clarify fully the role of the self-confidence stereotype dimension within this study: The fuzziness of the results for this dimension indicates that some more conceptual work needs to be done to catch all stereotypical content appropriately and exclusively.

In summary, our analyses produced some results that confirmed our assumptions but also some that were fairly surprising. Support was given to the collective and individual efficacy beliefs at the *Gymnasium*. The more positively teachers see their students stereotypically, the stronger are their beliefs that they can support them both individually and together with their colleagues. These effects are stronger in 5th grade when there is less real diagnostic information available to teachers than in 9th grade. For the *Hauptschule* an equivalent pattern shows for one of the humorous dimensions of collective expectancy beliefs. Another unexpected result was that the self-confidence stereotype showed a negative effect in 5th grade for individual as well as collective expectancy beliefs. A possible explanation of these surprising results is the fact that there are connotations within the

items that indicate stolidity. As the critically low reliability of the scale indicates, the items did not really measure the same thing. Another surprising finding was that in 9th grade at the *Hauptschule* the self-confidence stereotype dimension is associated with a lowered perception of obstacles in the classroom. Again, it would seem that there are connotations within the items that refer to a positive sense of self-confidence as well, at least in the case of the higher grade. These results indicate that there is some need for the clarification of stereotype content in order to develop instruments that are able to clearly catch stereotypic thoughts of teachers about their school track.

4.3 Strengths and limitations

The study at hand has substantial strengths: It is the first to show that teachers' stereotypes regarding the school type they teach at are connected to their efficacy beliefs. This is a very important result that should inspire new research on Pygmalion effects, this time with a focus on the collective dimension of teacher judgements. The present study is then the first step towards a comprehensive model of school stereotypes, whereby tracks represent only one feature that can be subject to stereotypes. Also individual classes and schools could be represented in a stereotypical manner. We were able to show substantial effects of the school track stereotypes on teacher efficacy beliefs despite the control for relevant student variables at school level. These results are stronger in 5th grade than in 9th grade, which indicates, in accordance with our assumption, that stereotypes are especially important where real prior experience to judge from is scarce.

Nevertheless, our study also has important limitations that should be borne in mind when interpreting the results. The major point is perhaps that we only conducted cross-sectional analyses. Although the PARS database contains longitudinal data, this was unfortunately not the case for the constructs included in the present study. Clearly, research designs with a longitudinal perspective are needed to prove the assumed causal directions. Also, the control for student variables should be applied within a multilevel approach at class level. In the study at hand it was limited by the fact that we could only control for student data at school level and through dummy variables at that. We had to apply this type of control at school level as class level identification variables that match teacher and student data are not available in the PARS database. Dummies were used as the number of schools was not high enough to apply multilevel modeling or treat the school data as an interval metric. Due to the low number of cases at school level, treating the data as metric variables may have biased the results due to the rather large distances between the data points. Thus, we decided to transform the data to ordinal level and used regular linear regression models. As our main purpose was to find out about the covariation of constructs at the teacher level, our study can be considered to yield robust estimates. Still, future studies should consider a multilevel modelling investigation of the relation of student development at Level 1 and teacher expect-

tancies as predictors at Level 2. In such research, student data should be considered at class level applying metric aggregations of the students' capabilities and cultural capital. The class level student data should be linked to the beliefs of the respective class teachers. Another drawback of our study is the exclusive use of extreme groups. A more comprehensive data set of the continuum of school types would make analyses of between schools more feasible.

As concerns the measures of the teachers' stereotypes, two limitations need to be highlighted. First, we only considered teacher statements on what they believed that people think about students of the *Gymnasium* and the *Hauptschule*. Our reason was that we assumed that this lowers potential biases due to answers considering social expectancies. Still, in further studies it would be better to carry out additional assessments of the explicit personal beliefs of the teachers. Although our expectation would be that both private and public stereotypical judgement dimensions would be strongly related, each covers different aspects of the stereotypical judgements about school types, which could lead to differential effects on behaviour and behaviour-relevant constructs like self-efficacy. Second, the two dimensions *humorous personality* and *self-confidence* need modification for the *Hauptschule*. The *self-confidence* scale especially has very low reliability that is below the usual acceptable levels. Accordingly, the results regarding this scale should be seen as no more than exploratory. But as low reliability usually leads to lower power and the results in this dimension are rather interesting, we decided to leave the scale within our models. It is all the more important to consider these results with great care as they might be due to random distribution because of the low reliability of the measure. Moreover, the reliabilities and the covariation effects identified indicate that further investigations are needed to develop a more comprehensive model of teacher stereotypes about the *Hauptschule*. We only applied a rather deductive stereotypical model, where items had been developed in an earlier study (Knigge, 2009) and were based on students' assumptions. Future work should apply a more inductive and theory-driven approach to define a more appropriate model for teacher stereotypes of school tracks. Further, the theory of teacher stereotypes should be elaborated: It is feasible that stereotypes about the two groups differ in their subcomponents. A consequence of our methodology was that we preferred yielding such distinct structures. This was, however, no problem for our investigation as we had planned to run regression analyses for both school types separately in any case. In this way, contextual influences were kept fairly constant and covariations of expectancies and stereotypes were captured more accurately. But in future, comprehensive theoretical models should be developed and supported with path and structural equation models. All in all, our study confirmed our belief that collective Pygmalion effects affect teacher behaviour. In the following last section, we will make a few brief suggestions on some desirable features of a research framework for such collective Pygmalion effects.

4.4 A short proposal of a research framework for collective Pygmalion effects

In the theory section of this article, it was proposed that in addition to individual student judgements teachers judge classes and schools as a whole. Cognitive resources are too limited to always process all students as individuals although a great deal of interaction goes on between teachers and their students as a group. The assumption was made that in addition to individual level judgements, judgements about the whole class would have a strong impact on teachers' behaviour in the classroom. Thus, these group judgements and resulting expectancies could lead to collective Pygmalion effects mediated through an adaptation of the teachers' behaviour to their beliefs in the collective dimension. Our study shows that it is worth investing some effort in the investigation of teacher expectancies and Pygmalion effects at the levels of classes as well as schools.

School type related stereotypes in the present study are, however, not considered to be the core of teacher expectancies in the collective dimension. It is of course necessary for such collective-level judgements to include also schemes for all the classes teachers teach and have taught. In addition, there should be schemes representing all the schools that they know. Besides one's own primary experiences, it is also secondary experiences, for instance from stories told by colleagues, that should influence judgements in the collective dimension and accordingly should have the potential to influence teacher behaviour in the classroom. Thus it is assumed that teacher judgements at least include experience-based and stereotype-based individual level schemes for individual students and experience-based and stereotype-based schemes for classes and schools, and finally also stereotypes regarding school tracks. But the latter should also be further elaborated as stereotypes exist not only with regard to explicit between-school tracking, but also about implicit forms of tracking and within-school tracking. In the light of the present results it seems reasonable to investigate such stereotypical content dimensions as well. To achieve a comprehensive model of teacher judgement effects in the collective dimension, it is necessary to consider implicit and within-school forms of tracking because in many countries only these forms of tracking exist.

In summary, the present article is to be considered only the first of many future studies. We are looking forward to the research that will evolve from this contribution and believe that there are a great number of other interesting aspects to discover in the realm of collective teacher judgments.

Acknowledgements

The authors would like to thank two anonymous editor's readers for their perceptive comments and helpful suggestions, and are also grateful to Michael Pätzold for valuable linguistic assistance with the article.

References

- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmileus. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 95–188). Wiesbaden, Germany: VS. doi: 10.1007/978-3-531-90082-7_4
- Bos, W., Lankes, E. M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (Eds.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster, Germany: Waxmann.
- Bos, W., Schwippert, K., & Stubbe, T. (2007). Die Kopplung von sozialer Herkunft und Schülerleistung im internationalen Vergleich. In W. Bos, S. Hornberg, 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., Stubbe, T. C., Buddeberg, M., Dohe, C., Kasper, D., Müller, S., & Walzebug, A. (2016). *Framework for the Panel Study at the Research School ‚Education and Capabilities‘ in North Rhine-Westphalia (PARS)*. Manuscript in preparation.
- Bortz, J., & Lienert, G. A. (1998). *Kurzgefaßte Statistik für die klinische Forschung. Ein praktischer Leitfaden für die Analyse kleiner Stichproben*. Berlin, Germany: Springer. doi: 10.1007/978-3-662-22078-8
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital. In R. Kreckel (Ed.), *Soziale Ungleichheiten. Soziale Welt* (Vol. 2, pp. 184–198). Göttingen, Germany: Otto Schartz & Co.
- Brophy, J. E. (1985). Teacher-Student interaction. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 303–328). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3rd, renewed and extended ed.). München, Germany: Pearson Studium.
- Dion, K. L. (2003). Prejudice, racism, and discrimination. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology* (Vol. 5, pp. 507–536). New York, NY: John Wiley & Sons. doi: 10.1002/0471264385.weio521
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behaviour: The mode model as an integrative framework. *Advances in experimental social psychology*, 23, 75–109. doi: 10.1016/S0065-2601(08)60318-4
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 357–411). New York, NY: McGraw-Hill.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. doi: 10.1037//0022-3514.82.6.878
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206–213. doi: 10.1007/s11121-007-0070-9
- Hamilton, D. L. (Ed.). (2005). *Social cognition: Key readings in social psychology*. New York, NY: Psychology Press.
- Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London, United Kingdom: Routledge.
- Heller, K. A., & Perleth, C. (2000). *KFT 4-12+R. Kognitiver Fähigkeitstest für 4. bis 12. Klasse, Revision*. Göttingen, Germany: Hogrefe.
- Helmke, A., & Jäger, R. S. (Eds.). (2002). *Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau, Germany: Verlag Empirische Pädagogik.

- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology, 101*(3), 662–670. doi: 10.1037/a0014306
- Jonsson, J. O. (1999). Explaining sex differences in educational choice – An empirical assessment of a rational choice model. *European Sociological Review, 15*(4), 391–404. doi: <http://dx.doi.org/10.1093/oxfordjournals.esr.a018272>
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131–155. doi: 10.1207/s15327957pspr0902_3
- Knigge, M. (2009). *Hauptschüler als Bildungsverlierer? Eine Studie zu Stigma und selbstbezogenem Wissen bei einer gesellschaftlichen Problemgruppe*. Münster, Germany: Waxmann.
- Knigge, M., & Hannover, B. (2011). Collective school-type identity: Predicting students' motivation beyond academic self-concept. *International Journal of Psychology, 46*(3), 191–205. doi: 10.1080/00207594.2010.529907
- Luedtke, O., Robitzsch, A., Trautwein, U., & Koeller, O. (2007). Umgang mit fehlenden Werten der psychologischen Forschung: Probleme und Lösungen [Handling of missing data in psychological research. Problems and solutions]. *Psychologische Rundschau, 58*(2), 103–117. doi: 10.1026/0033-3042.58.2.103
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2009). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106. doi: 10.1111/j.1750-8606.2008.00048.x
- McKown, C., & Weinstein, R. S. (2002). Modelling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology, 32*(1), 159–184. doi: 10.1111/j.1559-1816.2002.tb01425.x
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Opdenakker, M.-C., & van Damme, J. V. (2006). Differences between secondary schools: A study about school context, group composition, school practise, and school effects with special attention to public and catholic schools and type of schools. *School Effectiveness and School Improvement, 17*(1), 87–117. doi: 10.1080/09243450500264457
- Postmes, T., & Branscombe, N. R. (2010). *Rediscovering social identity: Core sources*. New York, NY: Psychology Press.
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85–97. doi: 10.1037/0022-0663.76.1.85
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom. Teacher expectation and pupils' intellectual development*. New York, NY: Holt, Rinehart & Winston. doi: 10.1002/1520-6807(196904)6:2<212::AID-PITS2310060223>3.0.CO;2-U
- Rubie-Davies, C. (2006). Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools, 43*(5), 537–552. doi: 10.1002/pits.20169
- Rubie-Davies, C. (2007). Classroom interactions: Exploring the practices of high and low expectation teachers. *British Journal of Educational Psychology, 77*(2), 289–306. doi: 10.1348/000709906x101601
- Rubie-Davies, C. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology, 80*(1), 121–135. doi: 10.1348/000709909X466334
- Schwarzer, R., & Jerusalem, M. (Eds.). (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin, Germany: Freie Universität Berlin.

- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. doi: 10.1002/mpr.329
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-Value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. doi: doi:10.1006/ceps.1999.1015