



Who buys into curricular reforms and why? Investigating predictors of reform ratings from teachers in Germany

Nicolas Hübner, Corey Savage, Cornelia Gräsel & Albrecht Wacker

To cite this article: Nicolas Hübner, Corey Savage, Cornelia Gräsel & Albrecht Wacker (2021) Who buys into curricular reforms and why? Investigating predictors of reform ratings from teachers in Germany, Journal of Curriculum Studies, 53:6, 802-820, DOI: [10.1080/00220272.2020.1870714](https://doi.org/10.1080/00220272.2020.1870714)

To link to this article: <https://doi.org/10.1080/00220272.2020.1870714>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 10 Mar 2021.



[Submit your article to this journal](#)



Article views: 2232



[View related articles](#)




[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

Who buys into curricular reforms and why? Investigating predictors of reform ratings from teachers in Germany

Nicolas Hübner ^a, Corey Savage ^a, Cornelia Gräsel^b and Albrecht Wacker^c

^aHector Research Institute of Education Sciences and Psychology, University of Tübingen, Germany; ^bInstitute of Educational Science, University of Wuppertal, Germany; ^cLudwigsburg University of Education, Germany

ABSTRACT

Prior research suggests that teachers' positive perceptions of a reform are key when it comes to its successful implementation. The importance of teachers as enactors of change efforts results from their close tie to what is happening in the classroom. This suggests that without persuaded change agents, innovations are unlikely to be implemented coherently, and in turn, likely to fail. In this study, we investigated the relationships between different sets of predictors (specific to the teacher, the school, the reform, or support structures) and teachers' overall rating of a curricular reform. We found that teachers' perceived added value and their perceptions of specific reform aspects predicted overall reform ratings. Furthermore, we identified heterogeneity regarding the importance of specific predictors between different school tracks, which calls for increased attention to school context when considering teachers' perceptions of reforms.




KEYWORDS

Curricular reform; policy implementation; teacher perceptions; school culture; school track

High achievement of students is seen as an important determinant of economic growth (Hanushek & Woessmann, 2008; Sahlberg, 2006); student life-course outcomes such as educational attainment, health, and earnings (e.g., Goldhaber & Özek, 2019) and the fostering of an informed and engaged citizenry (e.g., Galston, 2001), among other benefits. In recent years, many nations and educational systems have implemented educational reforms to improve student achievement (Hopmann, 2008; Organisation for Economic Co-operation and Development, 2015). In turn, schools all across the globe have been required to implement new curricula and modernized learning environments (Gerrard & Farrell, 2014; Hargreaves & Shirley, 2012; Hübner et al., 2020). In such times of global policy movements, knowledge on how to successfully implement intended changes is of great importance.

Teachers are the primary street-level bureaucrats of the education sector (Lipsky, 1980, 2010) and the 'ultimate enactors of any change effort' (Banner et al., 2012; R. E. Porter et al., 2015, p. 115). When implementing macro-level (i.e., system-wide) changes at the micro-level of the multilayered educational system (i.e., in the classroom), the relevance of teachers can be attributed to their immediate impact on student learning (Brühwiler & Blatchford, 2011).

Prior research suggests that teachers' buy-in of a reform can be understood as an antecedent of its successful implementation (e.g., A. C. Porter et al., 1988; Donnell & Gettinger, 2015; Fullan, 2015; Garvin & Roberto, 2005; Gräsel, 2010; Mayrowetz, 2009; Rogers, 2003). Furthermore, in early reform

CONTACT Nicolas Hübner  nicolas.huebner@uni-tuebingen.de  Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Europastraße 6, 72072 Tübingen, Germany
 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

stages, reform ratings can be used as proxies for successful future implementation and effectiveness (Carter, 2008; Donnell & Gettinger, 2015).

However, large-scale empirical studies that identify why teachers rate reforms more positively or negatively are lacking. In addition, minimal research exists on how school contextual factors, such as the school culture or learning environment, play a role in teacher reform ratings. The term school culture has a long history and has been used inconsistently in the past. Stoll (2000) gave a general definition on the foundations of school's cultures. She argues that a school's culture 'is shaped by the history, context, and the people in it' (p. 8). She further outlines that culture consists of several aspects related to (a) a school's age, (b) the external context of a school, (c) the type of a school, and (d) the student population at a school and their socioeconomic background. In this study, we considered three different groups of schools (school tracks), which vary in their school cultures including aspects such as teacher characteristics and their experiences (e.g., see Table 3), the student populations, and their official educational missions (e.g., preparing students for different diplomas).

In the case of curricular reforms, it seems very plausible that schools with different tracks, for instance, different student and teacher populations (e.g., Dockx et al., 2018; Klusmann et al., 2009), face different implementation challenges and concerns. The relevance of high fit between the school culture and reform requirements to the success of a reform has been explicitly theorized in earlier work (e.g., Hargreaves, 1997; Hargreaves & Shirley, 2012).

In the present study, we utilized a large dataset of teachers from lower secondary schools, which was collected during the second year of implementation of a statewide curricular reform. Based on previous research (Brezicha et al., 2015; A. C. Porter et al., 1988; Coburn, 2005; Donnell & Gettinger, 2015; Fullan, 2015; Gräsel, 2010; Prendergast & Treacy, 2018; Rogers, 2003), we identified relevant predictors (teacher-reported characteristics of themselves and their schools, teacher ratings of reform aspects and available support structures) and investigated their relation to teachers' global rating of the curricular reform. Following this, we investigated potential differences regarding the relation of these predictors with teachers' reform ratings across three different school types: Intermediate track schools (*Realschulen*), high track schools (*Gymnasien*), and comprehensive schools (*Gemeinschaftsschulen*). It is important to note that a range of comprehensive schools exist in Germany. In Baden-Württemberg, the comprehensive school (also referred to as community school) distinguishes itself from other schools by its focus on teaching at all achievement levels, its explicit focus on inclusive teaching, and mandatory all-day schooling at least 3 days per week (Bohl & Wacker, 2016). The results of our study provide crucial empirical evidence regarding key predictors of teacher buy-in, a core precursor for successful reform implementation and later reform success.

Education policy implementation and the relevance of teachers

As illustrated in our adapted educational policy implementation model (see Figure 1) compared to other fields of public policy, reforms in education can be characterized by a unique, field-specific implementation process (e.g., by means of professional development programmes or teacher

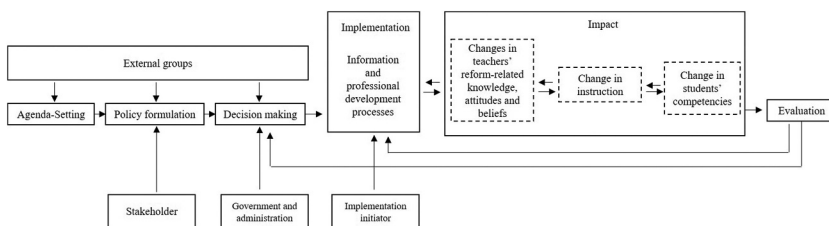


Figure 1. An adapted educational policy implementation model. Based on previous work by Lasswell (1956), Mayntz (1977), Jann and Wegrich (2007), and Desimone (2009).

training) and specific expected channels of impact (e.g., on teachers' attitudes, skills, and knowledge, classroom practices and subsequent student learning). As outlined, teachers are key for successful implementation (Banner et al., 2012; Lipsky, 1980, 2010), and their relevance results from their immediate impact on classroom processes and students' learning (Brühwiler & Blatchford, 2011). If reforms intend to result in changes on the micro-level (i.e., on student learning), these reforms must anticipate, monitor, and be closely aligned to the role of teachers in this process (Prendergast & Treacy, 2018). If teachers do not adapt and implement the intended changes, large-scale changes on student outcomes attributed to the reform are unlikely to occur (Donnell & Gettinger, 2015; R. E. Porter et al., 2015). More specifically, previous research has emphasized how teachers perceive a reform determines their engagement in appropriately implementing the required changes (Desimone, 2002; Easton & Erchul, 2011; Mayrowetz, 2009; Rogers, 2003). Practically speaking, if teachers' ratings regarding specific reforms are rather negative, the desired change is unlikely to happen (Achinstein & Ogawa, 2006; Choi & Walker, 2018; Terhart, 2013).

Although the importance of teacher buy-in for policy implementation has received attention through a theoretical and qualitative lens (e.g., Prendergast & Treacy, 2018; Wallace & Priestley, 2011), there is an absence of quantitative studies of why teachers buy into a reform, with few exceptions, to our knowledge. In one exception, Donnell and Gettinger (2015) investigated relations of belief congruence, self-efficacy, years of teaching experience, and professional development with the reform ratings of 209 elementary school teachers in southeastern Wisconsin during the implementation of Response to Intervention (RTI). This reform focused on implementing evidence-based instruction as well as intervention and prevention for students with academic or behavioural challenges. Their findings indicated that belief congruence, self-efficacy, and professional development were positively related to the acceptability of the reform. Interestingly, and in contrast to previous research (e.g., Hargreaves, 2005), years of teaching experience was not related to the rating of the reform. As is evident, the large-scale empirical evidence in this line of inquiry is very thin. As school reform efforts could be framed as reforms of teachers more than anything else (e.g., Darling-Hammond et al., 2017; Desimone, 2002, 2009; Leithwood et al., 2002; Thoonen et al., 2011), understanding the underlying mechanisms of teachers' reform ratings are crucial to more fully understand why teachers buy into reforms and why they do not.

School track and implementing policy reforms

While prior research has documented the importance of teachers for the successful implementation of any reform efforts (e.g., Desimone, 2002; R. E. Porter et al., 2015), school track or context is also an important consideration (Hargreaves & Shirley, 2012). As outlined in more detail below, teachers with varying qualifications and characteristics work in different types of schools with different student populations. In turn, why teachers rate a reform differently may vary across school tracks.

In tracked school systems (e.g., Germany), teachers in different school tracks or types typically have varying qualifications due to varying requirements for teacher education based on existing policy (e.g., Cortina & Thames, 2013). In addition to differences in formal qualifications, prior research in Germany also suggests that teachers in different school tracks/types differ on many additional characteristics such as secondary school grade point average, standardized achievement, general cognitive abilities, personality traits, and vocational interests (Klusmann et al., 2009; Roloff Henoch et al., 2015). Beyond Germany, explicit or implicit sorting of teachers with different qualifications and characteristics between and within schools is a global phenomenon (Kalogrides et al., 2013; Lankford et al., 2002; Luschei & Jeong, 2018). Teachers with different qualifications and/or characteristics may judge reform features and implementation barriers quite differently.

There also tend to be differences in student populations across different schools. In Germany, school tracks/types differ with regard to student achievement and a variety of other variables such as proportion of immigrant students and socioeconomic background (e.g., Guill et al., 2017; Stäbler et al., 2017). As shown by Dockx et al. (2018), such differences in student populations between school

tracks are also echoed in other countries like Belgium. In countries without explicit between-school tracking (like the United States), differences in student populations between schools are typically the result of socioeconomic segregation (e.g., Owens et al., 2016). Teachers who are confronted with different student populations may rate a reform more positively or negatively because of specificities of their respective student population.

The combination of different populations of teachers and students across schools may also require different adaptive behaviours for successful implementation. Other than general within- and between-school variation in implementation (Carroll et al., 2007), there may be different school cultural-related requirements regarding how teachers need to adapt to specific reforms in order to make them meaningful (Desimone, 2002). For instance, curricular reforms may aim to implement changes for specific groups of teachers or students, which are not equally distributed across all affected schools. Therefore, some teachers might rate aspects of the reform differently because they need to invest more effort to implement intended changes.

Overall, these findings suggest that different school tracks indeed provide different learning environments, as they are managed by different teachers and serve different students. These differences might be mirrored by differences in the way teachers are willing or able to handle the demands of curricular reforms or in the way they perceive the desired changes. In addition, these differences might result in differences in the ability and willingness to adopt policies and differences on the variables related to policy ratings (e.g., Van der Vegt et al., 2001).

The curricular reform

The curricular reform (Ministry of Education, Youth and Sports, 2016) focused on in this study was implemented in 2016 in the German state of Baden-Wuerttemberg. In Germany, the legal authority resides with the individual state (see, for instance, Education act of the State of Baden-Wuerttemberg §35–§37). Within each state, the ministry of education decides on the curriculum standards and the schools are expected to implement these standards.

As is visible in Table 1, the policy mainly focused on implementing four new global changes to the curriculum: (a) enhancing teaching based on educational standards, (b) the introduction of central 'guiding principles', (c) the introduction of three attainment levels in secondary school, and (d) the introduction of overarching educational foundations of democracy, peace, and cultural education (Ministry of Education, Youth and Sports, 2016). It is important to note that the implementation of the curriculum goes along with large amounts of autonomy by teachers. This is referred to in Germany as 'pedagogical responsibility' (Education Act of the State of Baden-Wuerttemberg §38), resembling the distinction between the intended curriculum and the implemented/enacted curriculum in international research on curriculum and opportunities to learn (Schmidt & Maier, 2009).

Regarding educational standards, content standards and process standards were defined, which captured both specific definitions of knowledge (acquired at a specific stage) and larger subject-specific dimensions of knowledge (acquired by the end of school). These standards were thought of as mandatory content, which must be taught in school by each teacher in a given subject and a given year. The definition and enhanced orientation of standards was also aimed at increasing the alignment between the curriculum and output-oriented governmental activities such as educational monitoring and accountability.

The introduction of central guiding principles was based on the idea of increasing student knowledge not only in specific subjects, but also in broader areas such as sustainable development or tolerance and diversity (see Concrete policy features in Table 1). These guiding principles were intended to be implemented across different subjects.

The adoption of three new attainment levels reflects the fact that learners in lower track schools (e.g., intermediate track or comprehensive schools) strongly vary from those in high track schools. Teachers in lower track schools are required to identify the different ability levels of students in class

Table 1. Key features of the final policy program and the implementation process.

Adopted policy feature	Concrete policy feature	Support structure	Intended aim	Track specific
Enhancing standards-based teaching	Definition of content standards (defining what students should know at each grade level) and process standards (defining subject-specific overarching competencies, which should be acquired by the end of education)		Enhance teaching based on educational standards. Enhance output-orientation governance	No
Definition of specific <i>guiding principles</i>	(1) Education for sustainable development (2) Education for tolerance and acceptance of diversity (3) Prevention and health promotion (4) Vocational orientation (5) Media education (6) Consumer education	A mixture of mandatory and optional professional development for teachers, online resources, and example curricula	Initiate discussions on and implement <i>guiding principles</i> in the curriculum of the school and the subject. To be taught across all subjects.	No
Introduction of <i>attainment levels</i> in secondary school	(1) Definition of three attainment levels (high, intermediate, low) (2) Operationalized by defining subject- and competence-specific operators (verbs describing teaching and learning behaviour)		Adaptive teaching on all three attainment levels (high, intermediate, low)	Only in non-high track secondary schools
Democratic, peace, and cultural education	Develop knowledge and skills of students in the fields of democratic, peace, and cultural education.		Develop knowledge and skills to let students contribute and value the three fields of education. To be taught across all subjects.	No

and provide more individualized instruction accordingly. In the high track schools, teachers are expected to teach only at the highest ability level.

Finally, as mentioned above, the new curriculum introduced general overarching educational foundations of democracy, peace, and cultural education. These should be taught across all subjects and are thought of as foundations of all educational processes at school (Ministry of Education, Youth and Sports, 2016).

Research questions

Based on previous research on educational reforms, which has underscored that teachers' positive perceptions and acceptability of reforms are key for successful implementation, we first investigated the relationships between teachers' global curricular reform rating and teacher characteristics, school characteristics, teacher perceptions of specific reform aspects, and the available support structures (Research Question 1). Next, we investigated the extent that the predictive power of these variables varied between different school tracks (Research Question 2).

We derived the different predictors based on prior research (e.g., A. C. Porter et al., 1988; Choi & Walker, 2018; Donnell & Gettinger, 2015; Fullan & Quinn, 2015; Gräsel, 2010; Rogers, 2003). To our knowledge, only a few promising, small-scale studies have investigated this topic empirically, and these studies have analysed only a subset of predictors from the theoretical and conceptual literature. In line with the available research, we assumed that teachers who report a greater added value of the reform and use more support structures should rate the curricular reform more positively. Regarding the added value, prior research has outlined that if the beliefs of the teacher

and the reform goals are more in line, they respond more positively to a reform (Donnell & Gettinger, 2015). Regarding the support structures, participation in professional development programs may be particularly fruitful. For instance, as outlined by Desimone (2009), successful professional development (PD) can help to shape teachers' beliefs, attitudes and knowledge of specific elements of content, all of which can help to build capacity to successfully implement reforms (e.g., Fusarelli, 2002). Theoretically, this line of argumentation is also in line with the policy attribute theory as explicated by A. C. Porter et al. (1988). As outlined by Desimone (2002), the more specific a reform is (e.g., in terms of PD) the more likely teachers buy in and implement reforms.

On the other hand, we assumed variables indicating higher demands for teachers to implement the reform (e.g., expected workload) to be negatively related to the curricular reform rating. This assumption results from literature on institutional capacity and successful policy implementation in school (e.g., Fusarelli, 2002; Malen et al., 2015), which has repeatedly outlined subjective demands of teachers as barriers to successful policy implementation (Durlak & DuPre, 2008). Regarding differences between school tracks, hypotheses are more difficult to specify and more exploratory in nature, given the dearth of existing evidence. We assumed that teachers from non-high track school types (i.e., intermediate track schools and comprehensive schools) might value support measures more, which are useful to implement the new curriculum. This results from the fact that teachers of non-high track school types had to cope with the new, highly demanding policy feature of teaching on different ability levels (see Table 1), whereas high track teachers only had to teach at one (the highest) ability level. In addition, because teachers from lower track schools are confronted with lower-performing students, they should also value the benefit of the new curriculum for their students more, compared to high track teachers.

Method

Sample

The data for this study were taken from an online survey (May–June 2018) of the aforementioned curricular reform, which was implemented in the school year starting in September 2016 in the state of Baden-Württemberg in Germany. Samples were drawn based on a probability proportional to size (PPS) sampling of schools, whereby larger schools had a greater probability to be drawn. This sampling procedure was chosen for its efficiency, as less schools are needed to achieve a large dataset of teachers. Overall, 20–22% of the population of schools per track were requested to participate in the study (60 intermediate track schools, 75 high track schools, and 60 comprehensive schools). The participation rate on the school level was high: 90% of the requested intermediate track schools, 83% of the high track schools, and 73% of the comprehensive schools participated in the survey. Due to the voluntary status of the survey and in line with previous comparable studies in this field (e.g., Wacker & Kramer, 2012), the participation rate on the teacher level was relatively low (intermediate track schools: 27%, high track schools: 29%, and comprehensive schools: 14%). Overall, we considered data of 1,132 teachers in three different school types (intermediate track: $n = 439$, high track: $n = 513$, comprehensive school: $n = 180$). In addition, data from 262 teachers with leadership duties were considered for estimating variables of the school environment.

School tracks

To be able to investigate differences between teachers from different tracks, we assessed teachers from different school tracks. The high track (academic school type; *Gymnasium*) is the typical learning environment for high-ability students in Germany and has a direct path to university. The intermediate (vocational) track schools (*Realschulen*) are the school type for students who show an intermediate school performance at the elementary level. Finally, the comprehensive schools (*Gemeinschaftsschulen*) display a mixture of all different tracks, much like comprehensive school

tracking in other educational systems such as in the United States. However, this school type is currently attended mostly by low to intermediately high performing students.

Measures

Teachers had to complete a single online survey in which they were asked to evaluate the curricular reform, its implementation at their school, as well as existing training and support measures. The completion of the survey took about 20 minutes. First, teachers entered basic information about themselves and their school, including, for example, their age, gender, and teaching load. Second, they were asked to evaluate the curricular reform. Third, teachers had to answer specific questions regarding how well different aspects of the reform worked. For example, teachers had to rate training measures, example curricula, and the implementation of the reform in their department. All questions and response options used in the survey are provided in [Table SA1 in Appendix A](#).

Global curricular reform rating

Global curricular reform rating, the outcome variable of this study, was directly measured by asking teachers for their overall rating of the curricular reform, ranging from very bad to very good on a 4-point Likert scale.

Teacher characteristics

Six variables were used to measure teacher characteristics including sex, age, teaching load, motivation, years of teaching experience, and number of previous curricula they had taught. As a measure of teaching load, participants had to indicate whether they had either a full or a reduced teaching load. They also had to indicate how the curricular reform affected their motivation.

Teacher ratings of reform aspects

Six variables were used to measure teacher ratings of reform aspects: added value for pupils, added value for teachers, value of the reform for the development of the education system, comprehensibility of the goals, and expected short-term and long-term workload to implement intended changes.

School characteristics

Teachers had to indicate whether they teach at an intermediate track school, a comprehensive school, or a high track school. Apart from that, seven variables were used to assess characteristics of the participants' schools. School size was measured as a categorical variable in terms of size of the student body. Two questions asked about the implementation of the curricular reform at their school. First, participants had to indicate whether their department had an implementation plan. Second, they had to indicate their school's main efforts to implement the reform. Teachers could indicate if their school developed an implementation plan (on department level or school conference level), had in-house school trainings, or utilized any further implementation efforts. Participants also filled in a nine-item scale regarding the cooperation of teachers at their school. Finally, the average rating of the curricular reform by personnel with leadership duties was estimated for each school as an indicator of the implementation and reform climate at the school.

External support structures

Six variables were used to assess the external support structures during the implementation process of the curricular reform. Three of those variables referred to the teacher professional development programme specifically developed for the implementation of the curriculum reform 2016. Participants rated both the quality and the sufficiency of the trainings and indicated how frequently they participated in such trainings. Participation frequency was measured as a categorical variable with four categories. The remaining three variables concerned the support measures that were

developed to help implement the curricular reform. First, participants filled in a six-item scale about how well they were informed about the various support measures. Second, there were two three-item scales, in which participants had to respectively rate the example curricula and the competency grid.

Statistical analysis

We first computed descriptive statistics for all variables related to the four theoretically derived sets of predictors (related to the teacher, the school, the reform and its implementation, and available support structures). Descriptive statistics are presented in [Table 2](#). We then specified five multiple linear regression models. In Model 1-Model 4 (M1-M4), we included a different, single set of predictors (from among the four sets of predictors) in each model, before we considered all predictors in a joint Model 5 (M5). All models included control variables of teachers' school type, gender, and age. The data preparation was conducted in R (R Development Core Team, 2019), and the final analyses were run in Mplus 8.0 (Muthén & Muthén, 1998–2017). All continuous variables were z-standardized separately before running the analysis, resulting in fully standardized regression coefficients.

Cluster structure

In order to account for the nesting of teachers in schools (dependency of teachers of the same school), we estimated cluster-robust standard errors (Snijders & Bosker, 2012) using a design-based correction method implemented in Mplus. Recent research has underscored the advantages of applying such adjustments over using more complex multilevel models (e.g., with random effects) to account for data nesting, due to fewer assumptions and straightforward interpretations (e.g., McNeish et al., 2017).

Missing data

A common challenge in social science is missing data, and various approaches exist to account for them. In this study, missing data were handled using full information maximum likelihood estimation (FIML), which was shown to be equally good or even superior to traditional approaches of dealing with missing data such as pairwise deletion or complete case analysis (e.g., Enders, 2010). To increase the plausibility of the missing at random assumption and the comparability of the models, we specified a common multivariate model and used variables not included as predictors in a specific model as auxiliary variables (Graham, 2003; Graham et al., 1997).

Results

Preliminary analysis

We first examined the descriptive statistics for the overall sample ([Table 2](#)) and for the three different school types ([Table 3](#)). Our overall sample consisted of 1,132 teachers, who on average were 41.36 years old and had been in the teaching profession for 12.71 years. Overall, 64% of the teachers who participated were female. The majority of teachers had already taught at least two curricula previously.

Regarding differences between school types (see [Table 3](#)), we found statistically significant differences on a broad variety of variables between teachers from different tracks.

Predictors of the global curricular reform rating (Research question 1)

Next, we had a closer look at the predictors of the global rating of the curricular reform by estimating stepwise multiple regression models. Results are displayed in [Table 4](#). In all models, we controlled for

Table 2. Descriptive statistics of the sample of the study.

	n	M	SD	Min	Max
Teacher characteristics					
Sex (1 = female)	1132	0.64	0.48	0	1
Age	1132	41.36	3.09	22	64
Teacher's motivation	801	2.23	0.72	1	4
Years teaching profession	1132	12.71	9.04	1	43
Teaching load (1 = reduced)	1132	0.42	0.49	0	1
Previously taught curricula	1132	2.44	0.72	1	5
Rating of reform aspects					
Added value students	1072	2.32	0.93	1	4
Added value teachers	1088	2.29	1.00	1	4
Comprehensibility goals	1079	2.86	0.85	1	4
Value for development	1011	2.24	0.99	1	4
Short-term workload	957	3.31	0.74	1	4
Long-term workload	870	3.03	0.69	1	4
School characteristics					
Rating: School leaders	977	2.74	0.50	1.33	4
Teacher cooperation	1114	2.55	0.59	1	4
Implementation plan	1092	0.69	0.46	0	1
Teachers' conference plan	1081	0.49	0.50	0	1
In-house trainings	1081	0.11	0.32	0	1
Further school-specific support structures	1081	0.08	0.26	0	1
School size	1121	3.97	1.27	1	6
Intermediate track	1132	0.39	0.49	0	1
Comprehensive school	1132	0.16	0.37	0	1
High track	1132	0.45	0.50	0	1
External support structures					
Training offer	951	2.49	0.85	1	4
Training quality	840	2.54	0.89	1	4
Training participation	1069	1.68	0.51	1	4
Support measures	1096	1.98	0.63	1	4
Rating example curricula	950	2.57	0.70	1	4
Rating competency grid	916	2.44	0.73	1	4
Dependent variable					
Global curricular reform rating	1030	2.56	0.55	1	4

Means and SDs were estimated using FIML.

gender, age, and the school type. In the first model, we considered variables related to the characteristics of the teachers. We found teachers' motivation ($\beta = .62, p < .001$) to be positively predictive of the rating of the curricular reform. Regarding the covariates, we found age to be negatively predictive of the rating of the reform ($\beta = -.13, p = .022$), suggesting that, holding constant the other variables in the model, older teachers generally rated the reform less favourably.

In the second model, we included variables related to the ratings of reform aspects, conditioning on the same covariates as the prior model. We found the rating of the added value for teachers ($\beta = .27, p < .001$) and students ($\beta = .15, p = .001$) to be statistically significantly related to the global curricular reform rating. In line with our hypothesis, the comprehensibility of the reform was positively predictive ($\beta = .21, p < .001$), suggesting that a higher rated comprehensibility, on average, went along with a better reform rating. Moreover, the value of the curricular reform for the development of the education system was positively related to the global rating ($\beta = .20, p < .001$). Finally, the expected long-term workload was negatively related to the global rating ($\beta = -.09, p = .001$).

In the third model, we had a closer look at variables related to characteristics of the school. Only the aggregated rating of teachers with leadership duties at the school had a positive relationship with individual teachers' global curricular reform rating ($\beta = .09, p = .018$). This implies that teachers

Table 3. Descriptive statistics for the three different educational tracks.

	High track (A)		Comprehensive school (B)		Intermediate track (C)		<i>p</i> _{AB}	<i>p</i> _{AC}	<i>p</i> _{BC}
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Teacher characteristics									
Sex (1 = female)	0.61	0.49	0.67	0.47	0.67	0.47	.124	.054	.969
Age	42.33	9.18	38.97	9.55	41.20	10.35	<.001	.146	.029
Teacher's motivation	2.41	0.70	2.24	0.74	2.02	0.69	.076	<.001	.023
Years teaching profession	12.94	7.97	10.43	8.78	13.38	10.12	.002	.550	.001
Teaching load (1 = reduced)	0.47	0.50	0.35	0.48	0.39	0.49	.017	.026	.479
Previously taught curricula	2.40	0.66	2.59	0.90	2.42	0.70	.016	.764	.038
Rating of reform aspects									
Added value students	2.52	0.93	2.53	0.94	2.00	0.85	.883	<.001	<.001
Added value teachers	2.54	1.00	2.57	1.00	1.88	0.87	.819	<.001	<.001
Comprehensibility goals	3.06	0.79	2.87	0.83	2.61	0.86	.017	<.001	.001
Value for development	2.52	0.96	2.45	1.00	1.81	0.86	.530	<.001	<.001
Short-term workload	3.29	0.66	3.31	0.78	3.34	0.81	.806	.286	.674
Long-term workload	2.88	0.64	3.09	0.77	3.19	0.67	.017	<.001	.268
School characteristics									
Rating: School leaders	2.93	0.36	2.84	0.45	2.48	0.54	.406	<.001	.004
Teacher cooperation	2.50	0.56	2.87	0.59	2.49	0.57	<.001	.850	<.001
Implementation plan	0.76	0.43	0.54	0.50	0.68	0.47	<.001	.032	.004
Teachers' conference plan	0.56	0.50	0.34	0.47	0.46	0.50	<.001	.009	.012
In-house trainings	0.07	0.25	0.08	0.27	0.18	0.39	.813	<.001	.004
Further school-specific support structures	0.08	0.28	0.13	0.33	0.05	0.21	.086	.034	.001
School size	4.70	1.09	2.98	1.10	3.52	1.03	<.001	<.001	.021
External support structures									
Training offer	2.73	0.83	2.27	0.74	2.31	0.84	<.001	<.001	.634
Training quality	2.86	0.82	2.35	0.87	2.25	0.88	<.001	<.001	.219
Training participation	1.72	0.48	1.55	0.49	1.68	0.54	.001	.265	.023
Support measures	1.97	0.65	2.16	0.62	1.92	0.60	.001	.340	<.001
Rating example curricula	2.69	0.68	2.61	0.73	2.43	0.68	.297	<.001	.035
Rating competency grid	2.40	0.74	2.59	0.78	2.41	0.70	.016	.811	.023
Dependent Variable									
Global curricular reform rating	2.72	0.67	2.70	0.75	2.33	0.75	.808	<.001	<.001

High track, *n* = 513; intermediate track, *n* = 439; comprehensive school, *n* = 180. Statistically significant *p*-values (*p* < .05) are printed bold.

at schools with more optimistic leadership personnel tend to rate the curricular reform more positively. Regarding the covariates, again age was found to be negatively related to the reform global rating. In addition, there was a statistically significant difference between teachers from high track schools and intermediate track schools ($\beta = -.50, p < .001$), suggesting that after accounting for the other variables in the model, intermediate track teachers rated the reform less positively compared to high track teachers.

In the fourth model, we examined variables related to external support structures. We found a positive coefficient for the quality of the professional development program on the global rating of the reform ($\beta = .20, p < .001$). The frequency of participation at the professional development program also had a positive, however somewhat smaller coefficient on the global curricular reform rating ($\beta = 0.06, p = .038$). The two support instruments, the example curricula and the competency grid, were both positively related to the overall reform rating: Teachers who rated the example curricula ($\beta = .22, p < .001$) and the competency grid ($\beta = .19, p < .001$) to be more helpful for implementing the curricular reform also reported a higher global rating of the reform.

Finally, in the fifth, overall model, the relations from the stepwise models remained statistically significant for a majority of variables. Regarding teacher characteristics, the rating of the effect of the

Table 4. Multiple regression results for predicting teachers' global curricular reform rating.

	M1		M2		M3		M4		M5	
	β	<i>p</i>	β	<i>p</i>	β	<i>p</i>	β	<i>p</i>	β	<i>p</i>
Teacher characteristics										
Sex (1 = female)	.07	.186	.10	.013	.05	.398	.02	.789	.06	.181
Age	-.13	.022	-.07	<.001	-.13	<.001	-.10	<.001	-.13	.002
Teacher's motivation	.62	<.001							.15	<.001
Years teaching profession	.06	.246							.07	.089
Teaching load (1 = reduced)	.00	.933							.03	.428
Previously taught curricula	.03	.358							.01	.787
Rating of reform aspects										
Added value students			.15	.001					.11	.012
Added value teacher			.27	<.001					.22	<.001
Comprehensibility goals			.21	<.001					.17	<.001
Value for development			.20	<.001					.17	<.001
Short-term workload			.01	.422					.02	.314
Long-term workload			-.09	.001					-.06	.020
School characteristics										
School leadership rating					.09	.018			.02	.248
Teacher cooperation					.05	.088			-.03	.248
Implementation plan					-.02	.578			-.03	.167
Teachers' conference plan					.04	.246			.01	.790
In-house trainings					.03	.457			-.01	.747
Further school-specific support structures					.05	.102			-.02	.314
School size					-.04	.319			-.01	.706
Intermediate track	-.24	<.001	.02	.667	-.50	<.001	.06	.537	.03	.633
Comprehensive school	.06	.354	.02	.657	-.14	.249	-.32	<.001	.08	.171
External support structures										
Training offer							.01	.828	.03	.302
Training quality							.20	<.001	.01	.723
Training participation							.06	.038	.05	.035
Support measures							-.02	.510	-.03	.095
Rating example curricula							.22	<.001	.05	.048
Rating competency grid							.19	<.001	.06	.036
R ²	.42		.61		.11		.32		.64	

All continuous coefficients are standardized. All estimations considered FIML to treat missing data. Analyses were based on an overall sample of $n = 1,132$ teachers. Statistically significant p -values ($p < .05$) are printed bold.

reform on teacher motivation remained positively associated with the global curricular reform rating ($\beta = .15, p < .001$). In addition, the age coefficient remained negative ($\beta = -.13, p = .002$), suggesting that older teachers, after accounting for all other variables in the model, generally rated the reform more negatively. Similarly, the coefficients for the ratings of reform aspects remained consistent in direction and significance with the results of Model 2. Added value for students ($\beta = .11, p = .012$) and teachers ($\beta = .22, p < .001$), the comprehensibility of the reform ($\beta = .17, p < .001$), and the value of the reform for the overall development of the education system ($\beta = .17, p < .001$) remained positive. The expected long-term workload of the new curriculum remained negatively related to the global curricular reform rating ($\beta = -.06, p = .020$), and expected short-term workload of the new curriculum remained not statistically significant. Regarding the school characteristics, we found no statistically significant relationship with global rating of the reform, after controlling for all other variables in the model. Finally, regarding the external support structures, the frequency of participation in professional development ($\beta = .05, p = .035$) and the ratings of the example curricula ($\beta = .05, p = .048$), and the competency grid support instrument ($\beta = .06, p = .036$) remained positively related to the global rating of the curricular reform.

Taken together, variables related to teachers' perceived added value for themselves, teacher motivation, the comprehensibility of the goals of the reform, the value for school development, and

Table 5. Multiple-group regression results for predicting teachers' global curricular reform rating: differences by educational track.

	High track (A)			Comprehensive school (B)			Intermediate track (C)			<i>p</i> _{AB}	<i>p</i> _{AC}	<i>p</i> _{BC}
	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>	β	<i>SE</i>	<i>p</i>			
Teacher characteristics												
Sex (1 = female)	.01	.06	.848	.13	.15	.372	.08	.06	.193	.435	.391	.749
Age	-.12	.07	.074	-.04	.09	.648	-.13	.06	.035	.493	.973	.447
Teacher's motivation	.10	.06	.110	.23	.07	.002	.18	.05	<.001	.160	.249	.631
Years teaching profession	.01	.07	.932	.06	.09	.494	.12	.05	.027	.633	.195	.587
Teaching load (1 = reduced)	.05	.06	.420	.00	.10	.997	-.01	.08	.871	.689	.528	.924
Previously taught curricula	.01	.04	.819	-.05	.06	.377	.03	.04	.570	.403	.800	.285
Rating of reform aspects												
Added value students	-.05	.05	.371	.01	.08	.866	.29	.05	<.001	.514	<.001	.006
Added value teachers	.23	.05	<.001	.41	.10	<.001	.14	.05	.005	.100	.180	.012
Comprehensibility goals	.22	.05	<.001	.12	.06	.046	.12	.05	.008	.197	.126	.990
Value for development	.34	.05	<.001	.17	.09	.074	.00	.04	.872	.087	<.001	.127
Short-term workload	-.00	.03	.910	.05	.06	.405	.04	.03	.197	.433	.347	.845
Long-term workload	-.07	.05	.120	-.01	.07	.862	-.07	.04	.093	.451	.916	.475
School characteristics												
School leadership rating	.04	.03	.186	-.03	.06	.652	.04	.02	.121	.336	.984	.321
Teacher cooperation	-.03	.04	.451	.00	.06	.938	-.07	.04	.074	.704	.507	.364
Implementation plan	-.01	.04	.707	-.17	.05	.001	.02	.03	.578	.008	.529	.002
Teachers' conference plan	.03	.04	.347	.01	.05	.823	-.02	.04	.525	.721	.278	.590
In-house trainings	.04	.04	.319	.02	.04	.609	-.07	.04	.062	.760	.044	.107
Further school-specific support structures	-.01	.04	.818	-.05	.04	.239	-.02	.03	.589	.449	.878	.547
School size	-.05	.02	.025	-.03	.05	.564	.05	.03	.079	.684	.006	.161
External support structures												
Training offer	.05	.04	.188	-.03	.07	.619	.00	.05	.929	.293	.431	.634
Training quality	.03	.04	.508	-.07	.09	.421	.02	.06	.718	.300	.900	.379
Training participation	.03	.04	.424	.07	.04	.069	.07	.04	.071	.435	.437	.985
Support measures	-.01	.03	.861	.00	.06	.983	-.06	.03	.061	.944	.247	.373
Rating example curricula	-.04	.04	.230	.11	.06	.058	.14	.04	.001	.027	.001	.635
Rating competency grid	.00	.04	.925	.02	.08	.763	.15	.05	.002	.824	.021	.152
R ²	.62			.71			.67					

High track, *n* = 513; intermediate track, *n* = 439; comprehensive school, *n* = 180. Statistically significant *p*-values (*p* < .05) are printed bold.

added value for students were most meaningful for the global judgement of the reform. Overall, the full model (Model 5) explained 64% of variance in teachers' global judgements of the reform.

Differences between different school tracks (Research question 2)

Next, we examined the predictive power of the independent variables by school type and tested for differences using a multiple group model of the full model (Model 5) displayed in Table 5. Regarding teacher characteristics, we found teachers' rating of the reform's impact on their motivation to be statistically significantly and positively related to the global rating of the reform at comprehensive schools ($\beta = .23, p = .002$) and intermediate track schools ($\beta = .18, p < .001$). The positive relation was not found to be significantly different between the three school types.

Regarding teacher ratings of reform aspects, we found the rating of the added value for teachers and the comprehensibility of the reform to be positively related to the global curricular reform rating in all school types (all *p* < .05). The relation of the added value for teachers was the highest in comprehensive schools ($\beta = .41, p < .001$) and was statistically significantly higher in comprehensive schools than in intermediate track schools ($\beta = .14, p = .005; p_{BC} = .012$). The positive relation between the comprehensibility of the reform goals and the global judgement of the reform did not

differ between tracks (all $p > .05$). Interestingly, the relation of the added value for students was statistically significant only in the intermediate track ($\beta = .29, p < .001$), and this relation was statistically significantly higher in the intermediate track than in both the high track and the comprehensive schools (track differences were $p \leq .006$). Finally, we found a statistically significant positive relationship between the value of the reform for the overall development of the education system and the global curricular reform rating only in high track schools ($\beta = .34, p < .001$), and this relation was statistically significantly higher in high track schools than in intermediate track schools ($\beta = .00, p = .872; p_{AC} < .001$).

Regarding school characteristics, we found a negative relation between the availability of an implementation plan at the school and the global curricular reform rating only for comprehensive schools ($\beta = -.17, p < .001$), whereas this variable was not statistically significant in high track ($\beta = -.01, p = .707$) and intermediate track schools ($\beta = .02, p = .587$). This relation differed significantly between the comprehensive schools and the two other tracks ($p_{AB} = .008$ and $p_{BC} = .002$). In addition, we estimated a negative coefficient for school size at high track schools ($\beta = -.05, p = .025$) and this coefficient was statistically significantly lower, compared to intermediate track schools ($\beta = .05, p = .079; p_{AC} = .006$), highlighting slightly lower overall ratings at larger, high track schools.

Finally, regarding the external support structures, the ratings of the example curricula and the competency grid were found to positively influence the reform rating only at the intermediate track schools ($\beta = .14, p = .001$ and $\beta = .15, p = .002$, respectively), whereas these ratings were not statistically significant for the two other tracks (all $p > .05$). These differences by school type were statistically significant when comparing the intermediate track with the high track (example curricula: $p_{AC} = .001$; competency grid: $p_{AC} = .021$), but not with comprehensive schools (all $p > .05$).

Taken together, the results suggest that, different school types indeed seem to provide different school cultures and although teachers at these different school types seem to value some things comparably in relation to the global rating of the reform, they also differ on some measures. For instance, the perceived added value to teachers was most relevant at comprehensive schools and less important at other school types. In contrast, the comprehensibility of the reform was equally relevant for teachers independent of the school type.

Discussion

In this study, we investigated the associations between teachers' overall rating of a curricular reform and teacher-reported characteristics of themselves and their schools, as well as teacher ratings of reform aspects, and available support structures. In particular, teachers' perceived added value for themselves, the comprehensibility of the goals of the reform, its perceived value for school development, teacher motivation, and the added value for students were found to be positively related to the global curricular reform rating, even after controlling for a large set of further variables. This finding is largely in line with suggestions from previous conceptual and theoretical research (e.g., Fullan & Quinn, 2015; Garvin & Roberto, 2005; Rogers, 2003).

Other findings were also quite interesting. We found a statistically significant negative relationship between teachers' age and the global judgement of the reform, which suggests that older teachers were generally rating the reform less favourably. This holds true even after accounting for the impact of the other variables in the different models and suggests that younger teachers may be more likely to buy into reform efforts, which would be in line with earlier findings (Hargreaves, 2005).

Furthermore, the participation in reform-specific professional development programs as well as the rating of the quality of the support instruments (example curricula and competency grid) were positively related to the global rating of the reform. This generally underscores the importance of accompanying reforms with adequate professional development programs, as suggested in previous research (Choi & Walker, 2018; Crockett, 2007; Donnell & Gettinger, 2015; Fishman et al.,

2003). Above and beyond professional development, supporting materials were found to be positively valued by teachers, likely because they help to implement the reform. From a larger perspective, these results are also in line with most prominent professional development theories, which suggest that such programs can foster changes in teachers' knowledge, beliefs, and attitudes (Borko et al., 2010; Darling-Hammond et al., 2017; Desimone, 2009). Such professional development programs may therefore weaken implementation barriers and foster processes of policy implementation through the development of teacher buy-in, in addition to other desired teacher outcomes.

Having a closer look at the predictive power of the different variables between different learning environments (i.e., school types), which has received very little attention in empirical studies in recent reform research, revealed another interesting finding: The way variables were related to global curricular reform ratings varied between different school types. For instance, the relationship between perceived added value for students and the global rating of the curriculum reform was substantially stronger at intermediate track (vocational) schools compared to both high track (academic) and comprehensive schools. Furthermore, the relationship between support structures and the overall reform judgement was stronger for intermediate track schools compared to high track schools. This finding is in line with research on implementation fidelity and the theory of culture and school change (Hargreaves, 1997; Hargreaves & Shirley, 2012), which underscores the need to more closely investigate school culture moderators of policy implementation (e.g., Carroll et al., 2007; Kurki et al., 2006). More specifically, our results suggest that in the case of the curricular reform 2016, one size did not fit all school types, and we were able to give insights into which variables might be differentially related to the global curricular reform rating between different school types. How our findings can be generalized to other reforms and school types or systems remains an important and open research question. Generally speaking, given the lack of existing empirical evidence, it is highly important to replicate our findings and test external validity in different reform settings, using different variables and considering data of other educational contexts. Given most educational systems tend to have some mechanism of sorting students and teachers to different schools, whether explicit or implicit, we expect similar findings beyond our sample and outside of Germany, but this needs to be tested in future research.

An important aspect to further investigate are subject-specific challenges and differences in the implementation of the reform and the global reform judgements. Up to now, very little quantitative research exists on how teachers translate overarching, global reform goals to a subject-specific level. Future research should therefore more explicitly consider subject-specificities and translation processes during times of change. In the case of our study, teachers were provided with example-curricula, which should help them to implement the required changes in class. Our results suggest that these subject-specific support measures were considered to be useful, at least in terms of the global reform judgement. However, future research is needed, which more broadly examines subject-specific implementation processes. For instance, while we focused on teacher perceptions of the curricular reform (the intended curriculum) to measure buy-in, future research in this line of inquiry should also consider what is actually taught in the classroom (the implemented curriculum) when considering buy-in and implementation fidelity.

Limitations

There are various strengths and limitations of this study, which should be kept in mind when interpreting the results. First, we were not able to implement a mandatory participation design; thus, the sample was based on voluntary participation of the teachers. This was due to considerations regarding high costs and potential response biases when forcing all teachers of the drawn schools to participate. Therefore, it remains somewhat unclear to what extent the findings of this study are generalizable to the whole population of teachers in the state. Unfortunately,

administrative teacher data were limited, so we were not able to compare sample characteristics to a large set of population characteristics. We were able to consider age, gender and school type from administrative data. Comparing these characteristics of the teachers in our dataset to population data revealed only small differences.

In the same vein, it is also important to note that the different groups of schools considered in this study have a unique history and culture, specific for the state of Baden-Württemberg. Generalization of findings therefore requires carefully comparing the culture and history of school types in this and other contexts. In addition, the curriculum is not entirely defined by the ministry of education. Its implementation in class substantially depends on the professional learning communities of teachers at a school. Due to school autonomy policies, which were implemented over the past decades, every school has to define its own curriculum to a certain extent, which might consist of varying foci and therefore likely varies across schools. In addition, different subjects will find different solutions on how to implement the requirements of the new reform in their unique subject-specific curriculum for specific grades. This has to be considered, when interpreting variation in the perception of new curricula by the teacher.

Second, the present study was based on cross-sectional data on the current reform only. Therefore, we were not able to investigate the relationship between variables conditional on ratings of the previous curriculum. Such data might have been particularly useful in order to investigate how teacher perceptions and judgements of earlier curricula or reforms impact global ratings of the reform under study and should be collected in future research.

Finally, as in most research, we were not able to consider all potentially relevant variables. What would have been highly interesting are links between teachers' stages of concern and the different predictors of global curricular reform ratings. The stages of concern questionnaire can assess teachers' concerns about reforms. Knowledge about teachers' concerns is important in order to adequately support them during change processes (e.g., George et al., 2008). In addition, as was shown in previous research, teachers between school types differ on a variety of variables (Klusmann et al., 2009), but we were not able to collect all of these measures in the present study (e.g., personality). Nevertheless, despite these potential limitations, our study provides an important quantitative extension of previous, mostly conceptual studies on the implementation process from the perspective of teachers.

Implications for policy and practice

Keeping in mind the potential limitations of this study, our study has three primary implications relevant for policy and practice. First, the perceived added value of the reform for teachers, comprehensibility of the reform, and its value for school development were key predictors of the global rating of the reform. Teachers who reported a higher added value for themselves, better comprehensibility of the reform, and a higher value for school development were, on average, more in favour of the reform. The results therefore underscore the relevance of a high fit between teacher perceptions and intended changes, which was often discussed as key to successful reform implementation (Donnell & Gettinger, 2015; Easton & Erchul, 2011).

Second, the availability and rating of support structures were found to play an important role for the global curricular reform rating. This is also in line with previous research, which has highlighted the importance of professional development for successful reform implementation (Choi & Walker, 2018; Fishman et al., 2003) and suggests that reformers should ensure that offers for professional development and other support structures are sufficient and useful for teachers. More specifically, professional development might be particularly useful in helping teachers fully understand demands and comprehend goals of a reform, which might also have a positive influence on the perceived added value. Professional development may also help to build support structures in terms of professional learning networks of teachers, which were discussed as helpful drivers of policy implementation (e.g., Lieberman, 2000). For policy, it seems particularly fruitful to consider more recent research from the state of Baden-Württemberg, which closely examines questions regarding the supply structure and corresponding

teacher perceptions of professional development in the respective state (e.g., Cramer et al., 2019). This could also be a great starting point for further investigations to increase knowledge about the effectiveness of the offers specifically created for training teachers to better implement the reform. Worth mentioning, without a suitable and coherent professionalization concept, professional development is unlikely to have major impacts on teachers in Germany. Given prior publications on the status quo of the professional development program in the respective state of Baden-Württemberg (e.g., Cramer et al., 2019), it seems that the supply structure did not stipulate professional development programs for all teachers in order to successfully implement the reform on a general and subject-specific level. Against this background, future research should more closely investigate the coherence of reform programs, which seem particularly important for successful implementation (Fullan & Quinn, 2015)

Finally, as the relevance of the different predictors to overall rating of the reform varied between different school types, different school tracks might perceive the challenges of policy implementation very differently. These findings call for both a deeper investigation of this line of inquiry and, potentially, the development of school-specific, tailored policy programs. Rather than overarching, one-size-fits-all programs, a more specific support and implementation structure, which fits the specificities of the culture of schools, may be needed for successful implementation of curricular reforms.

Acknowledgements

We would like to thank Tim Mori for his support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research project was supported by the Tübingen Postdoctoral Academy for Research on Education (PACE) at the Hector Research Institute of Education Sciences and Psychology, Tübingen, Germany; PACE is funded by the Baden-Württemberg Ministry of Science, Research and the Arts, Germany.

Notes on contributors

Nicolas Hübner is a postdoctoral researcher at the Hector Research Institute of Education Sciences and Psychology at the University of Tübingen. His current research focuses on the implementation and effects of educational reforms, curriculum research, teacher professional development, and student achievement and motivation.

Corey Savage is a senior research associate at the Annenberg Institute for School Reform at Brown University. He studies teacher policy and issues of equity and effectiveness in civic education.

Cornelia Gräsel is Professor of Educational Science at the School of Education at the University of Wuppertal. Her research focusses on cooperation of teachers, school innovations, and education for sustainable development.

Albrecht Wacker is Professor of School Pedagogics at the Institute of Education Sciences at Ludwigsburg University of Education. His research focuses on the implementation of policies in secondary schools and curriculum studies.

ORCID

Nicolas Hübner  <http://orcid.org/0000-0003-3528-8086>

Corey Savage  <http://orcid.org/0000-0003-3708-2196>

References

- Achinstein, B., & Ogawa, R. (2006). (In)Fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, 76(1), 30–63. <https://doi.org/10.17763/haer.76.1.e14543458r811864>
- Banner, I., Donnelly, J., & Ryder, J. (2012). Policy networks and boundary objects: Enacting curriculum reform in the absence of consensus. *Journal of Curriculum Studies*, 44(5), 577–598. <https://doi.org/10.1080/00220272.2012.699558>
- Bohl, T., & Wacker, A. (Eds.) (2016). *Die Einführung der Gemeinschaftsschule in Baden-Württemberg: Abschlussbericht der Wissenschaftlichen Begleitforschung (WissGem)* [The implementation of the comprehensive school in Baden-Württemberg: Final report of the scientific evaluation]. Waxmann Verlag.
- Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker, & B. McGraw, Eds. *International encyclopedia of education*. (3rd ed., pp. 548–556). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00654-0>
- Brezicha, K., Bergmark, U., & Mitra, D. L. (2015). One size does not fit all. *Educational Administration Quarterly*, 51(1), 96–132. <https://doi.org/10.1177/0013161X14521632>
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95–108. <https://doi.org/10.1016/j.learninstruc.2009.11.004>
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(40), 1–9. <https://doi.org/10.1186/1748-5908-2-40>
- Carter, S. L. (2008). A distributive model of treatment acceptability. *Education and Training in Developmental Disabilities*, 43(4), 411–420.
- Choi, T.-H., & Walker, A. D. (2018). A heuristic model for tailoring teacher development to educational reforms: Focusing on ambiguity and conflict generation. *Teaching and Teacher Education*, 74, 72–84. <https://doi.org/10.1016/j.tate.2018.04.013>
- Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19(3), 476–509. <https://doi.org/10.1177/0895904805276143>
- Cortina, K. S., & Thames, M. H. (2013). Teacher education in Germany. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (Vol. 14, pp. 49–62). Springer. https://doi.org/10.1007/978-1-4614-5149-5_3
- Cramer, C., Johannmeyer, K., & Drahmman, M. (2019). *Fortbildungen von Lehrerinnen und Lehrern in Baden-Württemberg. [Teacher professional development in Baden-Württemberg]*. GO Druck Media GmbH & Co. KG.
- Crockett, M. D. (2007). OP-ED: Teacher professional development as a critical resource in school reform. *Journal of Curriculum Studies*, 39(3), 253–263. <https://doi.org/10.1080/00220270701287867>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- Desimone, L. M. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479. <https://doi.org/10.3102/00346543072003433>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Dockx, J., De Fraine, B., & Vandecandelaere, M. (2018). Does the track matter? A comparison of students' achievement in different tracks. *Journal of Educational Psychology*, 111(5), 827–846. <https://doi.org/10.1037/edu0000305>
- Donnell, L. A., & Gettinger, M. (2015). Elementary school teachers' acceptability of school reform: Contribution of belief congruence, self-efficacy, and professional development. *Teaching and Teacher Education*, 51, 47–57. <https://doi.org/10.1016/j.tate.2015.06.003>
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>
- Easton, J. E., & Erchul, W. P. (2011). An exploration of teacher acceptability of treatment plan implementation: Monitoring and feedback methods. *Journal of Educational and Psychological Consultation*, 21(1), 56–77. <https://doi.org/10.1080/10474412.2011.544949>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Fishman, B., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, 19(6), 643–658. [https://doi.org/10.1016/S0742-051X\(03\)00059-3](https://doi.org/10.1016/S0742-051X(03)00059-3)
- Fullan, M. (2015). *The new meaning of educational change*. Teachers College Press.
- Fullan, M., & Quinn, J. (2015). *Coherence: The right drivers in action for schools, districts, and systems*. Corwin Press.
- Fusarelli, L. D. (2002). Tightly coupled policy in loosely coupled systems: Institutional capacity and organizational change. *Journal of Educational Administration*, 40(6), 561–575. <https://doi.org/10.1108/09578230210446045>
- Galston, W. A. (2001). Political knowledge, political engagement, and civic education. *Annual Review of Political Science*, 4(1), 217–234. <https://doi.org/10.1146/annurev.polisci.4.1.217>

- Garvin, D. A., & Roberto, M. (2005). Change through persuasion. *Harvard Business Review*, 83(2), 104–112.
- George, A. A., Hall, G. E., & Stiegelbauer, S. M. (2008). *Measuring implementation in schools: The stages of concern questionnaire*. Southwest Educational Development Laboratory.
- Gerrard, J., & Farrell, L. (2014). Remaking the professional teacher: Authority and curriculum reform. *Journal of Curriculum Studies*, 46(5), 634–655. <https://doi.org/10.1080/00220272.2013.854410>
- Goldhaber, D., & Özek, U. (2019). How much should we rely on student test achievement as a measure of success? *Educational Researcher*, 48(7), 479–483. <https://doi.org/10.3102/0013189X19874061>
- Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325–366). American Psychological Association.
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift Für Erziehungswissenschaft*, 13(1), 7–20. <https://doi.org/10.1007/s11618-010-0109-8>
- Guill, K., Lüdtke, O., & Köller, O. (2017). Academic tracking is related to gains in students' intelligence over four years: Evidence from a propensity score matching study. *Learning and Instruction*, 47, 43–52. <https://doi.org/10.1016/j.learninstruc.2016.10.001>
- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668. <https://doi.org/10.1257/jel.46.3.607>
- Hargreaves, A. (1997). Cultures of teaching and educational change. In B. J. Biddle, T. L. Good, & I. Goodson (Eds.), *Springer international handbooks of education: Vol. 3. International handbook of teachers and teaching* (Vol. 3, pp. 1297–1319). Springer. https://doi.org/10.1007/978-94-011-4942-6_31
- Hargreaves, A. (2005). Educational change takes ages: Life, career and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education*, 21(8), 967–983. <https://doi.org/10.1016/j.tate.2005.06.007>
- Hargreaves, A., & Shirley, D. (2012). *The global fourth way: The quest for educational excellence*. Corwin Press.
- Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417–456. <https://doi.org/10.1080/00220270801989818>
- Hübner, N., Wagner, W., Hochweber, J., Neumann, M., & Nagengast, B. (2020). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades! *Journal of Educational Psychology*, 112(1), 204–220. <https://doi.org/10.1037/edu0000351>
- Jann, W., & Wegrich, K. (2007). Theories of the policy cycle. In F. Fischer, G. Miller, & M. S. Sidney (Eds.), *Public administration and public policy: Vol. 125. Handbook of public policy analysis: Theory, politics, and methods* (pp. 43–62). Boca Raton: CRC/Taylor & Francis.
- Kalogrides, D., Loeb, S., & Bêteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, 86(2), 103–123. <https://doi.org/10.1177/0038040712456555>
- Klusmann, U., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2009). Eingangsvoraussetzungen beim Studienbeginn [Cognitive and psychosocial characteristics upon entry into teacher training]. *Zeitschrift Für Pädagogische Psychologie*, 23(3–4), 265–278. <https://doi.org/10.1024/10.1024/10.0652.23.34.265>
- Kurki, A., Boyle, A., & Aladjem, D. K. (2006). Implementation: Measuring and explaining the fidelity of CSR implementation. *Journal of Education for Students Placed at Risk (JESPAR)*, 11(3–4), 255–277. https://doi.org/10.1207/s10824669espr110304_3
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62. <https://doi.org/10.3102/01623737024001037>
- Lasswell, H. D. (1956). *The decision process: Seven categories of functional analysis*. College Park: University of Maryland Press.
- Leithwood, K., Jantzi, D., & Mascal, B. (2002). A framework for research on large-scale reform. *Journal of Educational Change*, 3(1), 7–33. <https://doi.org/10.1023/A:1016527421742>
- Lieberman, A. (2000). Networks as learning communities: Shaping the future of teacher development. *Journal of Teacher Education*, 51(3), 221–227. <https://doi.org/10.1177/0022487100051003010>
- Lipsky, M. (1980). Street-level bureaucracy: Dilemmas of the individual in public services. *Politics & Society*, 10(1), 116. <https://doi.org/10.1177/003232928001000113>
- Lipsky, M. (2010). Street-level bureaucracy: Dilemmas of the individual in public services. *The American Political Science Review*, 76(1), 275. <https://doi.org/10.2307/1960475>
- Luschei, T. F., & Jeong, D. W. (2018). Is teacher sorting a global phenomenon? Cross-national evidence on the nature and correlates of teacher quality opportunity gaps. *Educational Researcher*, 47(9), 556–576. <https://doi.org/10.3102/0013189X18794401>
- Malen, B., Rice, J. K., Matlach, L. K. B., Bowsher, A., Hoyer, K. M., & Hyde, L. H. (2015). Developing organizational capacity for implementing complex education reform initiatives. *Educational Administration Quarterly*, 51(1), 133–176. <https://doi.org/10.1177/0013161X14522482>

- Mayntz, R. (1977). Die Implementation politischer Programme: Theoretische Überlegungen zu einem neuen Forschungsgebiet [The implementation of political programs: Theoretical considerations on a new field of research]. *Die Verwaltung: Zeitschrift für Verwaltungsrecht und Verwaltungswissenschaften*, 10(1), 51–66.
- Mayrowetz, D. (2009). Instructional practice in the context of converging policies. Teaching mathematics in inclusive elementary classrooms in the standards reform era. *Educational Policy*, 23(4), 554–588. <https://doi.org/10.1177/0895904807312472>
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114–140. <https://doi.org/10.1037/met0000078>
- Ministry of Education, Youth and Sports. (2016). *Educational plan 2016*. Neckar-Verlag GmbH.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide*. (Eighth Ed.). Muthén & Muthén.
- Organisation for Economic Co-operation and Development. (2015). *Education policy outlook 2015: Making reforms happen*. OECD Publishing. <https://doi.org/10.1787/9789264225442-en>
- Owens, A., Reardon, S. F., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal*, 53(4), 1159–1197. <https://doi.org/10.3102/0002831216652722>
- Porter, A. C., Floden, R., Freeman, D., Schmidt, W., & Schwillie, J. (1988). Content determinants in elementary school mathematics. In D. Grouws & T. Cooney (Eds.), *Perspectives on research on effective mathematics teaching*. (pp. 96–113). National Council of Teachers of Mathematics.
- Porter, R. E., Fusarelli, L. D., & Fusarelli, B. C. (2015). Implementing the common core: How educators interpret curriculum reform. *Educational Policy*, 29(1), 111–139. <https://doi.org/10.1177/0895904814559248>
- Prendergast, M., & Treacy, P. (2018). Curriculum reform in Irish secondary schools – A focus on algebra. *Journal of Curriculum Studies*, 50(1), 126–143. <https://doi.org/10.1080/00220272.2017.1313315>
- R Development Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org>
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
- Roloff Henoch, J., Klusmann, U., Lüdtke, O., & Trautwein, U. (2015). Who becomes a teacher? Challenging the “negative selection” hypothesis. *Learning and Instruction*, 36, 46–56. <https://doi.org/10.1016/j.learninstruc.2014.11.005>
- Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change*, 7(4), 259–287. <https://doi.org/10.1007/s10833-005-4884-6>
- Schmidt, W. H., & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 541–559). Routledge.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.
- Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology*, 109(2), 191–207. <https://doi.org/10.1037/edu0000135>
- Stoll, L. (2000). School culture. *Set: Research Information for Teachers*, 9–14. Advance online publication. <https://doi.org/10.18296/set.0805>
- Terhart, E. (2013). Teacher resistance against school reform: Reflecting an inconvenient truth. *School Leadership & Management*, 33(5), 486–500. <https://doi.org/10.1080/13632434.2013.793494>
- Thoonen, E. E. J., Slegers, P. J. C., Oort, F. J., Peetsma, T. T. D., & Geijssels, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly*, 47(3), 496–536. <https://doi.org/10.1177/0013161X11400185>
- Van der Vegt, R., Smyth, L. F., & Vandenberghe, R. (2001). Implementing educational policy at the school level – Organization dynamics and teacher concerns. *Journal of Educational Administration*, 39(1), 8–23. <https://doi.org/10.1108/09578230110366883>
- Wacker, A., & Kramer, J. (2012). Vergleichsarbeiten in Baden-Württemberg [Standard comparative testing in Baden-Württemberg]. *Zeitschrift Für Erziehungswissenschaft*, 15(4), 683–706. <https://doi.org/10.1007/s11618-012-0326-4>
- Wallace, C. S., & Priestley, M. (2011). Teacher beliefs and the mediation of curriculum innovation in Scotland: A socio-cultural perspective on professional development and change. *Journal of Curriculum Studies*, 43(3), 357–381. <https://doi.org/10.1080/00220272.2011.563447>