Making sense of evidence-based governance reforms: an exploratory analysis of teachers coping with the Austrian performance standard policy.

Herbert Altrichter, Manuela Gamsjäger, Christine Plaimauer, Eva Prammer-Semmler, Regina Steiner, Julia Zuber & Lisa Handschuh

To cite this article: Herbert Altrichter, Manuela Gamsjäger, Christine Plaimauer, Eva Prammer-Semmler, Regina Steiner, Julia Zuber & Lisa Handschuh (2020): Making sense of evidence-based governance reforms: an exploratory analysis of teachers coping with the Austrian performance standard policy., Leadership and Policy in Schools, DOI: 10.1080/15700763.2020.1791906

To link to this article: https://doi.org/10.1080/15700763.2020.1791906
Making sense of evidence-based governance reforms: an exploratory analysis of teachers coping with the Austrian performance standard policy.

Herbert Altrichter, a, Manuela Gamsjäger b, Christine Plaimauer b, Eva Prammer-Semmler b, Regina Steiner b, Julia Zuber c, and Lisa Handschuha

aLinz School of Education, Department of Educational Research, Johannes Kepler University Linz, Linz, Austria; bUniversity of Education Upper Austria, Linz, Austria

ABSTRACT
During recent years many European countries have modernized the governance of their education systems according to an ‘evidence-based model’ which, e.g., materialized in new school inspections and comparative performance assessment. Qualitative case study data of six primary and secondary schools is used to explore in-school processes of sensemaking and constructing consequences of the Austrian performance standard policy (which is taken as an exemplar for evidence-based reforms). Teachers’ understandings and actions are compared with the normative claims underlying this policy. Results show that only two of the five processes claimed to be effective for school improvement through performance standards are found in the data.

Introduction
During recent years many European countries have attempted to modernize the governance of their education systems. In Austria – and, very similarly, in the education systems of the German federal states – the results of the international student performance assessment programs PISA and TIMSS and the political and media debate in their wake provided essential impulses for governments to initiate changes and thereby show leadership in a context of a proclaimed crisis (see Tillmann et al., 2008). The major model for modernization was the so-called evidence-based governance strategy which essentially promises to make educational decisions ‘more rational’ on all levels. Decisions about educational development – both on the macro-level of educational politics and central administration and on the level of individual schools and classrooms – are to be based on the best available research knowledge and recent evaluative information (Brown et al., 2017) instead of the good faith, political predispositions, or routines of actors on the political or practice level.

In this paper, we will focus on the question how teachers and school leaders perceive and interpret such a policy and to what actions they feel stimulated. In the beginning, we discuss the model’s normative aspirations and, in particular, the processes or mechanisms by which the evidence-based governance strategy aspires to become effective for educational development. Secondly, we will take the Austrian performance standard policy as an example of such instruments. By using case study data of six Austrian primary and secondary schools, we will explore teachers’ and school leaders’ perceptions and reactions to this policy and discuss whether these processes which the reform aspires to stimulate on school level can be found in our specific cases.
Evidence-based governance – ‘program theory’ and research findings

Evidence-based governance – intentions and mechanisms

Evidence-based policies – or accountability policies – are a variant of New Public Management (NPM)-based policies in education governance (Ball et al., 2012; Harris & Herrington, 2006; Maroy & Pons, 2019). They have been actively pursued in the school systems of Austria and Germany as a reaction to the PISA shock of 2001 which indicated that these school systems performed much worse than expected (Altrichter, 2020). In response to a heated public debate, the authorities prepared new governance instruments which were until then alien to these systems, among them performance standards, standard-based comparative student assessment and data feedback (Maag Merki, 2016), school inspections and quality audits (Altrichter & Kemethofer, 2016; Döbert et al., 2008), national or state institutes for educational measurement, national or regional education reports (Rürup et al., 2016).

In recent years there has been increasing criticism of the evidence-based strategy in German-speaking countries (e.g., Bellmann, 2016). Epistemological critique challenged the assumption that there were general answers for improving specific contexts and complex interactions (Berliner, 2002; Biesta, 2007; Heinrich, 2018). Another line of criticism pointed to simplistic models of evidence transfer and use (Maritzen, 2018), and more generally to the ambivalent attitude toward professionalism (H. Altrichter, 2020). It was also argued that comparatively high investments (H. Fend, 2011, 2018) were required for establishing the instruments of evidence-based governance – investments that went into the infrastructure of education with no direct benefit for the educational processes in classrooms (Piezunka, 2019, p. 271).

In this paper, we want to focus on another line of debate: The new evidence-based governance instruments promised to improve schools more quickly and more focussed than before. However, how – by what features, processes, and conditions – do these instruments aspire to achieve this goal and do we see these processes in the practice of schools? Research on the effects and processes of new governance instruments is certainly far from consolidated. Many individual studies are focussing on specific features of the policies, without seeing them in their more comprehensive context. We agree to Husfeldt (2011) who claimed that more knowledge about the appropriation and use of these instruments on the respective levels is necessary and that knowledge about several sub-processes of these governance models must be integrated into a more comprehensive conceptual model.

In a project about European school inspections, Ehren et al. (2013) analyzed documents about five inspection systems to produce a ‘program theory’ (i.e. a conceptual model organizing the elements and processes by which a ‘program’ aims to become effective for its stated goals; Leeuw, 2003). They claimed that these inspection systems converged in assuming that three processes are functional for achieving their improvement goals:

1. Setting expectations: By its criteria, documents, procedures, and overall organization, the inspection communicates expectations which schools will attend to in their work and thereby work toward the improvement goals.
2. Accepting feedback: Through inspection feedback and reports, inspections communicate to schools on what type and field of improvement to focus and thereby stimulate improvement.
3. Action of stakeholders: Through the visibility of the inspection process, its criteria and its results, the school’s stakeholders are alerted to the improvement needs of the school and stimulate action by support or pressure.

These three sub-processes of inspection seem to be characteristic for evidence-based governance strategies in general (Altrichter & Maag Merki, 2016; Harris & Herrington, 2006): Evidence-based or accountability policies set expectations for the performance of the education system and communicate them more clearly than before. They produce evidence about whether or not expectations have been met by the practical operation of the system units. They feed back this
evidence to actors on all system levels to stimulate and orientate improvement efforts. They include stakeholders, by asking their opinion (e.g., in inspection visits), by actively communicating quality standards and performance results to them, and by encouraging them to react to the performance of individual schools (e.g., as shown by comparative tests or inspection reports) through ‘voice’ or ‘choice’.

In the further parts of this paper, the Austrian performance standard policy is taken as an example for discussing the assumptions about effective processes inherent in evidence-based governance models. We will explain this policy in the next section.

**The Austrian performance standard policy**

In Austria, a policy of performance standards has been introduced since 2008. Performance standards for the primary cycle of schooling (students of 10 years) in Math and German (language of instruction) and for the lower secondary cycle (students of 14 years) in Math, German and English (main foreign language) have been formulated in a language of student competencies. After a pilot phase with a limited number of schools and a wave of publications and workshops, national standard testing for secondary students has been administered in 2012 for the first time and performance results have been fed back to schools (i.e. to students, teachers, parents, and administrators; see Altrichter & Kanape-Willingshofer, 2012). In 2013 the first round of national standard testing took place for primary schools. A support structure consisting of information brochures, teaching material, diagnostic tests, professional development courses, ‘feedback moderators’ (who explain the performance test results to schools), etc. has been gradually established.

Following a procedure by Leeuw (2003), Altrichter and Gamsjäger (2017) have analyzed legal and implementation documents in order to reconstruct the ‘conceptual model’ underpinning the Austrian performance standard policy. The goals of the policy are phrased as “broad normative fields” rather than a clear list of separate goals” (Altrichter & Gamsjäger, 2017, p. 12); they include improved student competencies, equity, and justice in education, and a range of process goals (such as developing a new, individualizing, result- and competence-oriented classroom culture, developing the quality of teaching and learning, school improvement) which are seen as conducive pathways to the main policy goals.

Furthermore, the model includes five processes which are meant to achieve the goals of improved performance and equity (see Figure 1) and which extend the model for school inspections by Ehren et al. (2013):

1. Setting expectations: By formulating and communicating performance standards more clearly than before, actors on all levels are provided with a clear reference system which will orient their actions and focus further improvement which, in consequence, will boost the quality of learning and results (see Figure 1, Line 1).
2. Stimulating by data feedback: The results of teaching and learning processes are measured by national comparative tests and fed back – in different aggregated versions – to students, teachers, school leaders, regional and central administrators. This data feedback will stimulate and orient the development of teaching and learning (see Figure 1, Line 3).
3. Alignment by support: A support structure will help actors (on classroom, school, and other levels) to translate the normative intentions of performance standards into concrete actions and structures (see Figure 1, line 2). Other support measures (such as brochures, feedback material, and ‘feedback moderators’) will help actors to make sense of data feedback and to derive practical consequences for improvement (see Figure 1, line 4).
4. Involving stakeholders: Schools are obliged to include parents (and in some secondary schools: student representatives) when they interpret data feedback reports and devise consequences for development. This will align in-school actors and school improvement to the ‘common good’ and the community’s expectations.
(5) *Alignment by in-school coordination*: Attention to the new performance standards requires intensified coordination between school management, subject faculties, and individual teachers which, in turn, will improve the quality of development processes and results.

This model is ‘normative’ in that it expresses the intentions of the Austrian legislators. However, the five processes of making the performance standard policy work are theoretically plausible (Ehren et al., 2015, p. 382): neo-institutionalist theories (Meyer & Rowan, 1977; Scott, 2001) can serve as justification for processes (1) and (3), social coordination theories (Schimank, 2007) substantiate processes (4) and (5), while process (2) is supported by feedback theories (Hattie & Timperley, 2007).

In this paper, we are interested whether these political aspirations and theoretical predictions are reflected by processes in schools. A policy and its implementation efforts are, in the first instance, ‘structural offers’ which have to be taken up and brought to life by actors on various system levels. This involves cognitive aspects of ‘sensemaking’ (noticing, interpreting, and constructing implications; Coburn & Turner, 2011), using existing contextual elements in a ‘bricolage’ manner (Maroy & Pons, 2019, p. 77), (micro-political) activities of negotiation, contestation or struggle between different groups (Ball et al., 2012), and processes of ‘re-contextualization’ (Fend, 2006) of reform elements to the language, structures, and practices on the various system levels (Spillane, 2012).

**Research on performance standards policies**

Performance standard policies are good examples of ‘traveling policies’ since variations of them have been implemented in many European education systems. However, when making comparisons between national education systems, Ozga & Jones’ (2006) warning should be kept in mind: the embedding of supposedly similar policies into different national contexts may lead to a wide variety of results. This has some impact on the discussion of research literature. The Austrian education system has been considered as a typical representative of a Central European bureaucratic and selective
configuration which is closer to the education systems of the German federal states than to Scandinavian and English systems (see Windzio et al., 2005). As a consequence, we will use international literature only for conceptual issues; however, when it comes to empirical questions, we will concentrate on research from German-speaking school systems.

To our knowledge, there are no studies that have comprehensively researched the processes of using the performance standard policy in Austrian schools. However, the implementation of performance standards in Austrian schools has been examined by a range of studies, mostly by questionnaires to teachers and principals (e.g., Freudenthaler & Specht, 2005, 2006; Grillitsch, 2010; see overview in Altrichter et al., 2016). The data are not in all respects comparable (due to changing sampling sizes and questionnaire wording), and the conditions of the pilot phase may differ from those of a standard phase. Nevertheless, some critical issues with respect to the sub-processes of making use of performance standard policies may be derived:

**Setting expectations**

Personal characteristics, such as acceptance, positive attitude and motivation to use the data, have been identified as key factors for the implementation of the standard reform (Ackeren, 2007; Groß Ophoff, 2013). More than half of those Austrian teachers who came into contact with performance standards during the pilot phase say that performance standards may be helpful for lesson planning, diagnoses of student learning, and self-reflection. However, the number of those who think that performance standards offer some additional orientation compared to the conventional syllabi is much lower (between 16% and 36%) and does not increase in the time observed (Freudenthaler & Specht, 2005, p. 21; Zuber, 2019). The group of teachers using standards regularly for lesson planning is even lower (between 2% and 14%) but increased up to 40% during nationwide implementation (Freudenthaler & Specht, 2005, p. 31, 2006, p. 17; Wacker, 2008; Zeitler et al., 2013; Zuber & Altrichter, 2018). Teachers are rather skeptical about the usefulness of performance standards for their classroom work; very few teachers (10% – 15%) expect that performance standards will make their work easier. There has been little change in teacher attitudes over the years: Strongly negative views of standards have become more neutral, but the number of positive attitudes is not increasing (Zuber, 2019). Comparative analyses indicate that the level of information, the extent of support, and the quantity of standard-related in-service training correlate with a more positive general assessment of performance standards, of their usefulness and with their actual use for lesson planning (Grillitsch, 2010, p. 99).

Case studies (Aiglsdorfer & Aigner, 2005; Hölzl & Rixinger, 2007) indicate that performance standards may have been communicated to students in a very restricted sense. Students in these schools seem to associate standards with testing and have not learned to use performance goals for organizing their learning. Lacking student participation may also be indicated by the fact that only 20% of students seek access to their performance results (Zuber et al., 2012).

If performance standards are understood as curricular elements that point to learning goals (Fend, 2008, p. 295) then these findings should not come as a surprise. We know from curriculum research (Wiater, 2006) that syllabi and curricular goals are quite slow instruments for reforming classroom work. On the other hand, many critics expected that performance standards will profoundly change classroom work through processes of ‘teaching to the test’ (TTTT). TTTT is usually considered an unintended or negative effect of comparative high-stake testing which will be detrimental to learning in several ways (Au, 2007; Deci et al., 1999, p. 633; Jäger et al., 2012): Content may be reduced to only those subjects which are tested. Learning processes may be fragmented by dissecting the subject in small proportions. Teacher control over learning may be increased. Learning time may be reallocated to tested subjects. Weak students may be excluded from learning opportunities to prevent them from sitting tests and decreasing the school’s average scores. However, other voices assume that TTTT may be conducive for learning as it focuses classroom work on relevant tasks and increases the motivation of students, parents, and professionals (Bishop, 1995, p. 658).

Due to the relative novelty of nation-wide comparative testing, there are few studies on TTTT in German-speaking education systems: Jäger et al. (2012) found that comparative performance tests –
even under low-stake conditions in two German states – are associated with narrowing the curriculum. Narrowing the curriculum is correlated with less certainty and lower self-efficacy of teachers and does not decrease over a longer period with more test experience (see Oerke et al., 2013). Maag Merki et al. (2010, p. 48) found both positive and negative effects of comparative low-stake performance testing on the quality of classroom instruction (as indicated by cognitive activation). On the whole, effects varied between education systems, subjects, and course types which did not allow for conclusive interpretations.

**Stimulating by data feedback**

Feedback is considered an effective mechanism of interpersonal adaption. An extensive body of research shows that it may have positive effects on learning; however, feedback cues, task characteristics and situational and personal variables may moderate the effect of feedback (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). There is some debate, whether it is feasible to transfer interpersonal feedback findings to the more complex conditions of institutionalized data feedback in a multi-level system (Altrichter et al., 2016, p. 247; Visscher & Coe, 2002).

Empirical studies on data feedback test results in Austria and German education systems indicate (Altrichter et al., 2016; Maier & Kuper, 2012): Many teachers (and even those teachers who say they are open toward data feedback and ready to analyze it) do not find it easy to process information about their students’ performance and derive practical consequences from it. Most studies found disappointingly little use of data feedback for classroom development. Data feedback seems to trigger some discussion in staffrooms; however, more substantial changes in classroom practice are rare (e.g., Grabensberger et al., 2008). Most studies found, if any, low-level consequences, such as repetition of content and tasks (Groß Ophoff et al., 2006; Maier, 2008), using test formats directly for teaching (Hosenfeld et al., 2007; Maier, 2008), minor adjustments of classroom interaction (see Schildkamp et al., 2009) or ‘symbolic use’ of data feedback in debates with parents and school management (Groß Ophoff, 2013, p. 295). If consequences are implemented in a classroom, they are usually not transferred to other classrooms (Groß Ophoff et al., 2007, p. 423). Also over the past ten years, the – already low – use of data feedback in schools has declined in some studies (Groß Ophoff et al., 2019).

Contrary to the limited use of external data feedback, teachers say that they base the development of their teaching on other types of data, mostly on self-generated data such as internal evaluations, assessment results or stakeholder interviews (Hult & Edström, 2016; Kemethofer et al., 2015). Teachers seem to be more ready to use data feedback for assessment and diagnosis rather than for developing teaching because they associate test data with the testing of individuals and not with classroom development (Maier, 2009).

Limited use of feedback data for classroom development does not seem to primarily depend on the technical or language quality of the feedback text. However, some presentation features can increase teachers’ motivation for data feedback reflection. The aggregation level of the reported results is particularly important: More profound data representation triggers individual diagnostic use of data and increases the willingness to reflect results (Breiter & Light, 2006). Including ‘further questions’ at the end of data feedback boosts teachers’ examination of the results, however, without significant influences on classroom development actions (Merk et al., 2018).

An important condition for using such data is an ongoing practice of continuous development and self-evaluation on the school level. In such a context, performance data may be a helpful source of information, among others. If this context is missing, performance data are an alien element or interpreted as a threat to established practices (Kuper, 2005).

**Alignment by support**

A support structure is considered as an essential and conducive factor for the implementation of reforms (Fullan & Stiegelbauer, 1991). The availability of supportive teaching material has been identified as one of the most important influences for the implementation of educational standard
reforms. Even in the case of negative attitudes toward change, teachers perceive good support material as helpful and use it for their teaching (Prenger & Schildkamp, 2018; Zuber, 2019). Professional development opportunities seem to be helpful: Grillitsch (2010, p. 99) found that the level of information, the extent of support and the quantity of in-service training correlate with more positive attitudes to performance standards and with the increased use of performance standards for lesson planning. Contrary to professional support, collegial and informal support has not proved conducive for enhancing data use (Schildkamp et al., 2014; Zuber, 2019; Zuber & Altrichter, 2018).

A central barrier to data use may be the understanding of the statistical information. Teachers seem to focus primarily on simple information, such as arithmetic means, and do not take in more sophisticated information which may give impulses for instructional development, such as effect sizes or mean differences (Merk et al., 2018). Support persons for reading and interpreting data feedback (‘feedback moderators’) were used by about a third of Austrian schools. School leaders found them useful for data interpretation, however, no further effects of this support measure on the staff’s interpretation of data feedback were found (Grabensberger et al., 2008; p. 70; Rieß & Zuber, 2014, p. 36; van der Scheer et al. 2017; Zuber & Altrichter, 2018). Scheer, Glas & Visscher (2017) conclude that a deeper use of data feedback can only be achieved through intensive teacher training interventions that go beyond explanatory support formats.

**Involving stakeholders**
Research on standard implementation tells us very little about the views and actions of actors other than in-school professionals. So far, the implementation of performance standards has not triggered increased parental involvement (Hölzl & Rixinger, 2007, p. 124; Nachtigall, 2005; Zuber & Katstaller, 2018). Even if schools prepare special information about standard testing and test results for parents, no additional motivation for parental participation in school development is found (Zuber & Katstaller, 2018).

**Alignment by in-school coordination**
The use of performance standards and data feedback is more likely in schools which have built up a culture of collaboration and mutual support in staff rooms (Asbrand et al., 2012; Gathen, 2011, p. 20; Pont et al., 2008, p. 5lf.). The implementation of performance standards may itself stimulate communication in staff rooms (Freudenthaler & Specht, 2006, p. 20). However, until now there are more reports on loose, informal communication rather than on long-term coordinated improvement work (Aiglsdorfer & Aigner, 2005, p. 225; Grabensberger et al., 2008). A key position for stimulating teachers’ use of performance standards and data feedback is attributed to school leaders (Hartung-Beck, 2009, p. 244; Musić et al., 2013).

**Research design**

**Research questions**
In sum, studies about the implementation of performance standard policies in German-speaking countries show fewer effects on school and teaching improvement than expected. With respect to the five intermediary processes which are meant to produce the proclaimed outcomes, there is little evidence that they are as powerful in Central European low-stake systems as the policy’s proponents suggest. For some processes (e.g., process 1, process 3, process 5) previous findings are ambivalent, and for others, they are negative (process 2). For process 4 very little research evidence is available.

Another characteristic of previous research is that it usually does not investigate the various performance standard processes in their connection and interaction. There is much research on data feedback and much less both on the effects of the performance standards on classroom teaching and on the role of external stakeholders. However, there is virtually none that tries to build a more
comprehensive perspective on these processes in their relationship, which would be essential for understanding the appropriation of the policy on school level.

In this exploratory study, we want to use qualitative case study data to learn more about how Austrian teachers and school leaders make sense of the educational standards reform and what actions they take when they aim to implement it on school and classroom level. In particular, we will focus on two research questions:

1. What measures of school and teaching improvement do teachers and school leaders of Austrian primary and secondary schools report in reaction to the implementation of the performance standards policy?
2. What intermediary processes (which are assumed to be effective for school and teaching development by the ‘conceptual model’ of the performance standard policy) are reflected in the perceptions and narratives of the actors at school level?

The overall idea is to understand what development processes are stimulated by the Austrian ‘performance standard policy’, which may serve as a first indication whether or not the implementation of this policy may achieve what it aspires to. In the empirical sections of this paper, we will ask teachers and school leaders what actions they have taken in the face of the several elements of the performance standards policy (e.g., standards, data feedback, stakeholder reactions). Research Question 1 focusses on reports about the ‘target activities’ of classroom and school improvement which are aiming for improved performance and equity. Research Question 2 takes a look at accounts about the ‘intermediary processes’ which are hypothesized by the conceptual model to be decisive stimuli for classroom and school development.

**Methods**

Previous research on performance standards mostly used quantitative questionnaire data (see Asbrand et al., 2012, for an exception). Thus, it seemed worthwhile to learn more about the processes of ‘recontextualizing’ the performance standard reform on school level by using a more open qualitative case-based approach. Data were collected through document analyses and qualitative guided interviews with teachers, school leaders, and representatives of students and parents in three Austrian primary and three secondary schools. Interview guidelines included a set of common questions which were to be followed up by more open invitations to narrate activities and processes in the wake of the policy reform. With respect to research question 1, reactions to direct questions for activities of classroom and school improvement (e.g., Has your teaching changed since the introduction of performance standards? If so, in what way? [Interview guide for teachers, 1st wave]; What decisions for teaching development have been taken in the staff meetings? [Interview guide for teachers, 2nd wave]) were analyzed. Concerning the ‘intermediary processes’ in research question 2, no direct questions were aiming at the five model processes. Rather teachers’ and school leaders’ narratives about individual and collective activities in the face of elements of the performance standard reform were interpretatively assigned to the five processes of the ‘conceptual model’.

Interviewing took place in a longitudinal design at three dates in the course of three consecutive years, i.e. (t1) in the school year before national testing of subject A (i.e. Math in primary and secondary schools), (t2) about three months after performance data feedback of subject A was given to schools and before national testing of subject B (i.e. German in primary and English in secondary schools) took place; (t3) about three to six months after performance data feedback of subject B was communicated to schools. Data were analyzed through qualitative content analysis (Kuckartz, 2012; Mayring, 2014); in a first step case studies on individual schools were written (Yin, 2013) which were subjected to a cross-case analysis (Feldman et al., 2018, p. 208) in a second step.
**Cases and sampling**

The argument of this paper is based on three primary and three secondary\(^2\) school case studies. All schools are located within the same large region (approx. 1.5 million inhabitants). They were selected through an analysis of school homepages and expert advice by regional administrators to offer variation in the working conditions according to the criterion ‘developmentally active school vs. less active school’.

Within every school approx. 12 – 18 interviews (depending on school size) with teachers, school leaders, and representatives of students and parents were conducted. Respondents were selected according to two criteria: On one hand, representatives of each of the following actor groups on school level were interviewed: the school leader, the faculty leader of subject A (tested in the first testing cycle), a teacher union representative, and a representative of both students and parents. In the first round of data collection, interviewees were asked for names of additional interview partners promising to offer additional and alternative perspectives on the school’s appropriation of performance standard policies who were interviewed in a second cycle of interviewing.

Due to these characteristics of methodology and sampling, this study does not aspire to come up with generalizable results; rather it should allow exploration of some long-term processes by which schools and actors make sense of, recontextualize, and react to performance standard policies. The expected results are hypotheses about trajectories of influence and possible effects of these policies on classroom and school development which may orient and stimulate further research.

**Results**

**School and teaching development**

Research question 1 asks the professionals in our case schools for measures of classroom and school improvement, which they have taken as a reaction to the performance standards policy.

**Changes in classroom teaching**

In all schools studied here, teachers in the tested classes report different types of changes in their classroom teaching. Firstly, teachers of both primary and secondary schools redesign worksheets and assessment tasks for *practicing the new test formats* (e.g., ‘multiple choice’) which have so far played little role in assessment with their students. These changes are meant as preparation for the standard assessment and follow the ‘teaching to the test’ strategy (Au, 2007).

Secondly, many teachers in secondary schools (see e.g., AF1/ML1t; SD2/ML2t\(^3\)) shift, emphasize or de-emphasize *lesson content*. E.g. ‘Statistics’ is no longer taught at the end of the 8th grade, but earlier, because this content is included in the mathematics standard test. Some content is taught ‘more intensively’ than before, e.g. Listening Comprehension in English (SD1/E1 t) or some aspects of mathematics in primary schools.

Thirdly, there are also cases when teachers – in line with the policy expectations – are reporting more profound, ‘competence-oriented’ changes in their teaching. Some (however, not all) teachers in all three primary schools say that they are now using a wider variety of tasks and that their classroom work increasingly focusses on meaningful reading or independent and logical thinking in Mathematics.

>[We] have frequently built in these tasks [from the Internet or from accompanying material] that actually promote logical thinking or common sense. (SE2/L1)

Similarly, some teachers (quite consistently in the secondary schools except for the academic secondary school) say that the ‘reasoning and arguing in maths’ or ‘listening comprehension and writing in English’ have become much more important in their teaching. Particularly in the secondary school E, concerted efforts toward competence-oriented teaching using competence grids seem to have been pursued in the faculty groups (see SE3/Dt; SE3/LD).
**School improvement**

In five of the six schools, no new *institutions of teacher collaboration* have been created (or existing ones have been changed) before the standard testing. Only in primary school B, the principal asks fourth-year teachers for joint curriculum planning. Some primary teachers say that performance standards stimulated their informal communication. In secondary school E, interviewees think that teamwork has increased in recent years, which is, however, attributed to a simultaneous secondary school reform rather than to the performance standard policy (e.g., SE3/Et). In the other two secondary schools, performance standards have not occupied a prominent role in collegial and professional collaboration (e.g., SD2/M1t). In the academic secondary school interviewees even deny that performance standards have been discussed in the faculty groups (AF2/M2t).

After the data feedback, there are three instances of improvement measures with the potential to go beyond individual classrooms. The principal of primary school B uses results from the data feedback and from diagnostic tests to schedule ‘development conversations’ with the respective teachers, which are meant to alert teachers to the diagnostic test results and “the actual standard testing, and discuss the consequences that each team draws from this.” (PB2/SL) The principal of secondary school D encourages the English faculty to develop ‘text writing’ strategies in the next school year (SD3/E2t; SD3/SL). In the academic secondary school, whole-school improvement is triggered by below-average school climate figures in the data feedback which the school principal wants to make a topic for development work.

**Development for equity aspects**

Our data do not indicate that teachers feel stimulated by the performance standard policy to launch special development activities concerning equity or to pay special attention to equity aspects in other improvement efforts. This seems to reflect the conceptual criticism that the equity goals of this reform are verbally proclaimed, however, only vaguely supported by structural provisions (Altrichter & Kanape-Willingshofer, 2012). On the contrary, teachers in primary school A (e.g., PA/L2) and in secondary schools D and E warn that competency-based classroom material provided by the authorities may be too difficult for weaker students and, thereby, denying them the experience of success.

**Mechanisms**

In research question 2 we asked whether those processes which are assumed to be effective for school and teaching development by the performance standard policy are reflected in the perceptions and narratives of the actors at school level.

**Setting expectations**

The reform’s intention to promote individualized and competence-oriented learning is supported by most interviewees from our primary schools and seems to be in line with their professional understanding. However, for many primary school teachers, this is not their main association with ‘performance standards’; rather they seem to associate ‘performance standards’ primarily with ‘standardization’ and ‘testing’ which they are very skeptical of.

... because, again, everyone is asked the same thing [in standard tests]. And when I try to support [special] skills of a child in individualized learning [...] and suddenly the child is assessed exactly the same as everyone else. Then that does not fit together for me. (PC3/L1)

Nevertheless, they feel obliged to prepare their students for these assessments by practicing new test formats and partly changing their teaching as has been explained in the previous section.

The situation is somewhat different in secondary schools. In none of our case schools, ‘performance standards’ are the main reform topic. The policy’s messages are superimposed by other simultaneous reforms, by the secondary school reform at schools D and E and by the new “centralized school-leaving examination (Zentralmatura)” at the academic secondary school F. However, the educational
implications of these reforms are seen to be in line with the demands of the performance standard policy by most respondents, e.g., the standard competency models are compatible with the competence grids which are used to facilitate development work in the secondary school reform (e.g., SE3/LB).

In sum, the standard reform has sparked expectations for changes in classroom teaching in the six schools studied, ranging from (more frequently) relatively ‘superficial’ adaptation processes to (less frequently) more coordinated and long-term development of competency-based learning. Many of these changes may be seen as instances of ‘teaching to the test’ which is usually associated with a deterioration in learning opportunities for students when teachers restrict their learning offerings to the content and test formats they expect in the exams (see Au, 2007; Jäger et al., 2012). Our interview material certainly offers limited opportunities to assess the quality of teaching changes; additional analysis of teaching materials and lesson observations would be more informative.

However, we believe that we have found both indications of limitations as well as of extensions of learning possibilities in the interview data: Having students practicing test formats will probably be regarded as an instance of restricting the demands and goals of mathematics education. On the other hand, reinforcing ‘reasoning and arguing in maths’ could well be an extension of learning opportunities for many students. A student representative in the academic secondary school of our study claims that more attention to the students’ understanding of central concepts was paid in those lessons meant to prepare for the comparative tests:

… in the chapters which we did [as preparation] for the testing, there was more focussed learning. […] Because … so the teacher explained it better than before. And it was also important to her that everyone really understood it. (AF2/SVt, 4)

The English teachers of secondary school E provide a particularly interesting case. In the interviews, they say that many of the changes (e.g., listening comprehension, writing texts) introduced in preparation for the comparative tests are reasonable and fully conform to their idea of modern English teaching methodology. They have learned this teaching methodology in their teacher education or in professional development, but they have apparently not consistently implemented it, because – as they explain – it makes testing more difficult or time-consuming, it is associated with organizational and technical difficulties or teaching materials were difficult to obtain. Most of these obstacles seem to be swept away (or at least alleviated) by the support materials offered by the performance standard implementation so that teachers feel ready for such changes in their teaching.

Why we have not done it [used the competence-oriented teaching strategies before] over and over, and why we have sometimes neglected it? […] is just that … because things have to be checked and free speech is not so easy to test – it is already clear, but it is very time-consuming. If I have a certain learning objective in mind, then I’m in a time constraint and then … yes, then I’ll just leave it or just shorten it. (SD1/E1t)

**Stimulating by data feedback**

A core element of the impact model of the performance standard policy is data feedback. Test results are communicated to relevant actors in aggregated form: the students receive their individual results, the teachers get the results aggregated at the class level, and the school management the results of all tested classes of their school. For a fair comparison, so-called ‘expectancy values’ are offered which account for a number of social indicators. The data feedback report displays whether the results of a school fall into an ‘expectancy range’, i.e. do not significantly differ from schools working under similar socio-economic conditions.

What processes can be observed in the case schools? About three to four months after data feedback, only a primary school teacher has specific results in some detail in mind, which she wants to use for improvement in the following school year (PC2/L1t). In both secondary schools, the mathematics results are no longer "known in detail". When asked about the data feedback, the principal in school E stands up to collect a special file from which he reads the main results (SE2/
SL). These observations suggest that the feedback results have not left deep traces in the school’s self-
interpretation.

In all six schools, the overall results are remembered as ‘positive and very satisfactory’ by the
interviewees. The ‘expectancy range’ is an essential reference point for the interpretation (SD3/E3). To
be within the expected range produces some sense of “confirmation”: the previous work was fine, and
no changes or further development are necessary (see AF2/ML2t; SE2/SL).

Only if results fall below the expectancy range in specific dimensions, further deliberations may
take place. E.g., the unsatisfactory results in geometry tasks are explained by a mathematics teacher
that these contents are located at the end of the textbook and have not been taught before testing (SD2/
MF2t). Other explanations are found in the students’ capabilities (e.g., SD3/E2 t; SE2/M2t) or the
specific contents of the standard test which had not been part of conventional teaching (SD2/E2t). The
mediocre results of the English test in the academic secondary school are attributed to the school’s
focus on natural science which disadvantages it in comparison to language-oriented grammar schools
(AF2/E2t).

What consequences are drawn from data feedback? The primary principals want to learn from
classes with good results. While the principal of school B invites the teachers of the tested classes for
a ‘developmental conversation’, those of school A and C insist that all learning will be voluntarily and
that they cannot enforce classroom changes or professional training (PC3/SL). In secondary school E,
which had been very active in the previous secondary school reform, consequences of standard-related
feedback are not an issue at all. In the academic secondary school, consequences for teaching are left to
individual teachers and are not discussed by the faculty groups. However, the principal energetically
embarks on improving the school climate (where the school has underperformed in the standard
feedback) and makes it the subject of school-wide development efforts.

Although the principal of secondary school D, who is a mathematics teacher himself, is quite
capable of deriving some ideas for teaching development from the data feedback, there is no concerted
interpretation of the results. Neither is the faculty group required by the school leader to construct
action consequences. There are some ideas for teaching development which are proposed by individ-
ual teachers; however, they remain ‘individual’ and are not coordinated at school or subject group
level. These innovatory ideas often opt for changing the content (“reinforcing geometry teaching”),
and more rarely for changing teaching and learning processes (similar results in Groß Ophoff et al.,
2006, p. 8; Maier, 2008). This may substantiate the assessment of one of the school leaders, who – in
the third round of interviews – does not expect a lasting effect of data feedback for teaching
development:

I think the teachers are perceiving the results, they give them a moment’s thought, but then they go back to
business as usual very quickly. (SD3/SL)

Alignment through support
The implementation of the standard policy is complemented by various support instruments, such as
information leaflets, professional training for school leaders, etc. However, the most relevant support
measures seem to be the following ones.

Diagnostic tests. In the course of implementation efforts, diagnostic tests fitting to the standard
competence model had been developed by the state institute. These tests allow testing students’
standard-oriented competencies in the years before the comparative standard test takes place.

The leaders of the primary schools A and B (similarly the head of the academic secondary school)
explicitly promote usage of this test in their schools as they consider it a useful tool for school
development and for feedback on classroom teaching. The diagnostic test is also appreciated by
many teachers, as it allows timely feedback on their teaching and the progress of student
performance.
Teaching material. The teaching material provided by the state institute is assessed by many teachers, in particular by primary school interviewees, as useful and supportive. Also, several teachers in the secondary schools D and F (e.g., SD3/LD; AF3/ArgeL) use this support material. It seems to offer an officially legitimized interpretation and concretization of what is meant by ‘performance standards’ and practical support for competence-oriented classroom teaching (see also Frühwacht, 2012, p. 168).

Yes, yes, in German, we have [this support material] in German, too, So that’s really good for [having students] practicing. You don’t have to reinvent everything, if there is something to support you. (PC1/SL, 2). However, teaching material may be perceived and used differently according to different professional frames of reference (Zeitler et al., 2013): The principal of primary school C associates ‘competence orientation’ with the ‘teaching and learning’, and attributes great potential to the material for lower-performing students and for individualized teaching. Other teachers whose prime association is ‘standardized testing’, use the material for test preparation and high-performance students (PA2/L2).

Feedback moderators. Schools may call in ‘feedback moderators’ if they look for expert support for analyzing and interpreting data feedback. According to their job description, they must not participate in developing improvement consequences but must limit themselves to data interpretation.

These support persons have actually been used by most schools of our study (except the academic secondary school). Sometimes the feedback sessions include some teachers of the tested subjects, most of the time it is just the principal who wants to substantiate his/her understanding of the feedback since s/he has to communicate it to teachers and parents. The informational quality and the interactional climate of these sessions usually receive positive feedback; however, they do not seem to trigger much practical improvement consequences (e.g., Grabensberger et al., 2008; Rieß & Zuber, 2014).

Involving stakeholders

According to a ministerial circular, each school must make a part of the data feedback report accessible to the parent and student representatives by a given date and must discuss it in a ‘school partnership meeting’. This obviously aims for ensuring that data feedback is not interpreted unilaterally from a professional perspective but from the broader perspectives of different school partners. Additionally, the results of the performance standard tests and the consequences to be drawn must be included in a ‘development plan’ which is discussed with the school’s regional administrator as a part of a new system-wide quality management system (Altrichter, 2017).

In our data, there is no reflection of the role of standard results in the ‘development plans’ and the conversations with regional administrators, however, this may be due to the novelty of the quality management model at the time of our interviews. All schools presented their results in school partnership meetings in the first year of the study (e.g., SE2/SL; SD2/SL; AF1/EV). In the second year, one of the principals postpones the presentation to a regular school partnership meeting which takes place – contrary to the ministerial guidelines – much later in the year. The design of these events is mostly information-centered: Test results are explained by the principal by a PowerPoint presentation, usually with no special arrangements for triggering parent or student reactions.

Some (not all) parents’ representatives and most in-school professionals think that parents are lacking interest in standard testing results because of two reasons: (1) Due to the long period between testing and data feedback parent and student representatives do not see their ‘own’ results, but those of the previous year (e.g., SD2/SL). (2) For parents of fourth or eighth graders, other topics are more important, i.e. parents in primary schools are more interested in the transfer to secondary education, those in secondary schools in further vocational and educational opportunities (SD1/SL), while those in academic secondary schools want to learn about the new centralized school-leaving examination system.

In sum, stakeholder involvement in the analysis of the performance results was not intensive and showed little to no impact on school and lesson improvement. A possible explanation (in addition to
those in the data) may be that stakeholder inclusion runs contrary to the bureaucratic history of Austrian schooling: Schools were seen for a long time as distant links in a bureaucratic chain; they derived their authority from their association with the state which provided these services for the public good. Parents and students were beneficiaries of these services but not meant to interfere with them. As a consequence, parents were traditionally viewed as ‘external persons’. Schools tried to keep distance to parents and ‘other outsiders’ to discourage their interference in in-school affairs.

Alignment by in-school coordination

Another assumption of the standard policy is that requirements for competence-oriented teaching and for shared interpretation of data feedback will strengthen in-school coordination. Reversely, tighter in-school coordination will strengthen individual teachers’ commitment to the reform goals.

Most schools in our data do not set up any new structures or processes to address challenges that might have been brought by performance standard policy. Only in primary school B, the school leader calls for reliable coordination meetings of year teachers (which is criticized as an unwarranted “march in lockstep” by some teachers; PB3/L1). Usually, the existing collaboration patterns are used for coping with the challenges of performance standards and are further strengthened by standard-related activities. E.g., the faculty groups in secondary school E, which had been very active before, meet four times a year and “examine standards for their implications for learning design” (SE3/Dt; SE3/LD). The English faculty of school D seems to collaborate similarly.

There is no evidence in our data that formal faculty meetings have taken place in secondary school D (except for English) and in the academic secondary school for coordinated preparation for standard-based instruction, for shared interpretation of the data feedback and for the planning of improvement consequences (AF2/M2t). It is interesting to see that the principals in all three secondary schools seem to be reluctant to launch projects of teaching development as a consequence of data feedback, although they have both an above-average understanding of the test results and ideas for adequate improvement measures. In these cases, the task of classroom improvement seems to be assigned to subject teachers and their professional judgment and capacities.

Different leadership concepts may underpin these observations: The school leader of primary school B is more resolute in her requirements for the teachers of the 4th grade; however, she also transfers rights of decision to the group of teachers. Other school leaders seem to more clearly attend to the traditional borders between teacher autonomy and school administration (PA1/SL, 3). E.g., the principal of the academic secondary schools does not interfere in classroom development, however, he is most active in promoting consequences on the whole-school level through a project of improving school climate.

“Well, I think it’s up to the school leader. If you take advantage of that [i.e. the demands of the performance standards policy] a little bit and make something out of it, it can be very beneficial [for school improvement]. It’s then again up to the school management how you deal with it. . . . . . . [She gives an example for her monitoring of the teachers’ work with data feedback:] I look at the comparison of the diagnostic test and then the actual standard testing, and at the consequences that the respective teams draw from it, how they work with it.” (PB2/SL, 5)

Summary, interpretation and further considerations

In qualitative case study approach we explored how teachers and school leaders in three Austrian secondary and three primary schools perceived and interpreted the challenges of a performance standard reform and by what improvement measures they reacted to it.

Research question 1: The performance standard policy seems to stimulate changes in the teaching of the subjects tested in our case schools. Changes include new assessment formats, shifts and re-balancing of contents; however, they also introduce competence goals which had not been given much attention before but are suggested by the performance standard policy. It is striking that in two schools these changes are mainly dealt with individually by the subject teachers or via informal
contacts between them, but not through existing structures (such as the subject faculty) or through coordinated initiatives pushed forward by school principals. ‘Whole-school development’ that goes beyond the teaching of individual teachers can be found in a few cases, e.g., a secondary school principal launches a project to improve the school climate as a consequence of standard feedback. The idea that performance standards can increase equity in education (which appears in official goal descriptions) plays hardly any role in the schools’ reasoning about performance standards. On the contrary, some teachers express their concern that the assessment tasks of standard testing may be too challenging for weak students.

In research question 2 the case data was checked for five intermediary processes proposed by a conceptual model of the impact of the performance standard policy: The teachers’ and school leaders’ accounts of perceptions and activities in the wake of the performance standard reform in our six case schools seem to indicate that this reform is, indeed, ‘setting expectations’ for teaching and learning. In some places, classroom improvement efforts are seen as a reasonable continuation of previous efforts of promoting individualized teaching, a secondary school reform or the centralized school-leaving tests. In other places, changes in classroom teaching seem to happen through the ‘Teaching to the test’ mechanism: Teachers try to prepare their students for the standard assessment by adapting the content of their teaching, the learning assignments and the assessment formats to the perceived demands of the performance standard policy.

The competence-oriented teaching material and diagnostic tests provided by the implementation support seem to be well received and have some potential for guiding the process of perceiving and implementing the performance standard reform, i.e. the intermediary process ‘alignment by support’ seems to be working in this case. This is not true for ‘feedback moderators’ who – though well-received on an interactional level – do not have any impact on improvement action.

The process of ‘receiving and interpreting data feedback’, on the other hand, has little influence on changes in teaching in the schools studied here. While individual school leaders and teachers can derive some plausible interpretations and ideas for development from the data feedback, there is little evidence in our material – apart from the climate initiative of the academic secondary school – that coordinated efforts of school and teaching improvement are stimulated by data-feedback. This may be disappointing for proponents of evidence-based governance since standard assessment and data feedback occupy a central role in their models of stimulating rational and well-focused improvement. In our longitudinal data, the teachers’ skepticism toward the external testing decreases, most likely due to the growing experience that no negative consequences are connected with the performance results. However, time does not seem to change the fact that the messages of the data feedback are not analyzed in detail. The schools’ assessment scores are checked to see if they are within (or above) the ‘expectancy range’. In the positive case, no further analysis of the information and no development measures are considered necessary. In the case of comparatively bad climate values in the academic secondary school, the principal becomes active. A ‘rationalization’ of teaching development through better information, as has been hoped in the evidence-based model, is scarcely seen in the cases examined. In general, the implementation of competence-based teaching, but above all, the processing of data feedback often seems to be delegated to individual teachers. The existing formats for collegial coordination are used for implementing the performance standard policy and seem to suffice for the professionals interviewed.

Probably due to the late date of data feedback, but perhaps also because of the traditional distance of Austrian schools to parents, ‘including stakeholders’ is no effective mechanism for implementing the performance standard reform.

The results of our exploratory study are certainly no valid basis for general claims, however, it is not unreasonable to assume that there will be a sub-group within the population of schools which will show similar reactions to performance standard policies and other evidence-based instruments. As a consequence, it is also possible to derive hypotheses for further studies. What are the messages of our case schools for the evidence-based governance model? First of all, it will be good to hear for the proponents of the standard reform that the professionals in these schools do not complain about
interference with other current reform projects. However, it is unfavorable for implementation that the school leaders in our schools (who mostly understand the reform principles and agree with them more than many teachers) show little initiative for initiating and encouraging coordinated development measures based on performance standards and data feedback; rather, they seem to categorize it as a ‘teaching innovation’ which is to be carried out by individual teachers into whose autonomous work they do not want to intervene.

If (as suggested by our cases) the implementation of these policy reform ideas works on the (traditional) way of ‘setting normative expectation’ rather than through the ‘more modern’ mechanism of ‘stimulating development through feedback’, the advocates of evidence-based governance will be disappointed. Standard testing and data feedback seem to have less impact on the development of classroom teaching than ‘impending tests’ and preparation for them. However, this result is not surprising and may be explained in analogy to results of school inspection research (Ehren et al., 2015) with neo-institutionalist arguments: Reasonably good and developmentally active schools react proactively in challenging situations and prepare for being tested. Then, the results will usually have no surprises for them and do not trigger systematic development. Rather ‘bad schools’ are characterized by difficulties with goal-oriented development before being tested. After test results have been made available, they cannot easily switch into a development mode. Rather, it is even more unlikely that productive development will take place because unfavorable data feedback has increased the pressure on the school. In situations of high pressure, it is more likely that a school will turn to quick-fix solutions, rather than to engage in sober data analysis and systematic planning of improvement strategies.

What conclusions could be drawn from our findings, if replicated in further studies, for the design of standards-based innovation processes? If the goal is to encourage schools and teachers to make use of performance feedback, it takes a lot more than just communicating standards and test results to schools. In particular, the role of school principals seems too vague in the implementation strategy and should be supported better in further implementation efforts. The setting of expectations and the investment in support processes currently appear to be more plausible ways of implementing the intentions of the performance standard policy.

What are the limitations of the present study? Apart from the small number of cases used in our exploratory approach, the number and quality of reform activities may be overestimated because of the exclusive use of interview data and self-reports. Additional observation studies would be desirable. Another limitation is that none of our schools received assessment results below the expectancy range. It is possible that more (or other types of) attention will be paid to the analysis of results and to further development measures in such cases. How schools cope with unfavorable results could be a worthwhile focus of further studies. It might also be argued that the implementation of a comprehensive innovation on classroom and school level is a long-term process lasting for more than three years.

At the level of the practical implementation, it became clear that the implementation of the ideas and instruments of the performance standard policy entails different processes and produces different results even in our small number of schools, very much depending on the existing collaboration culture. School principals, who in our schools were more likely to give positive opinions on performance standards, often seemed to shy away from encouraging or even demanding coordinated teacher action in the implementation of competence-oriented teaching or the use of the standard feedback. Performance standards seem to be understood as ‘teaching-related innovation’ that falls into the realm of teacher autonomy. In this respect, the performance standard policy is in a situation like many other innovations: It is intended to stimulate teaching and school development through its structure and its instruments. However, it only succeeds (as e.g., in secondary school E), where teaching and school development is already happening and where the school has previously built up structures of collaboration and of translating reform impulses into feasible work practices.
Notes

1. The analysis in this paper will concentrate on teacher and school leader data. Student and parent data is used in a few instances to complement the professionals’ perspectives.

2. For the secondary school case studies, two ‘Neue Mittelschulen’ (NMS) and an academic secondary school (Allgemeinbildende Höhere Schule; AHS) were chosen to represent the bipartite Austrian secondary system. After four years of primary schooling (6–10 years), two types of secondary schools cater for Austrian students: the Neue Mittelschule (NMS) represents the ‘practical’ and less prestigious track of the Austrian secondary education system and caters for students of 10 to 14 years. The Allgemeinbildende Höhere Schulen (AHS) represents the academic (‘Gymnasium’) track of the secondary system and caters for students from 10 to 18 years. In times of declining student numbers, there is often a fierce competition for students which is biased toward the more prestigious academic secondary schools. For distinction, we will call them ‘secondary school’ and ‘academic secondary school’ in this paper.

3. For the references to data see Appendix 1.

4. School improvement (as opposed to classroom development) is here defined as whole-school development which goes beyond changes in individual classroom teaching.

5. It is a peculiarity of the Austrian standard concept is that the results of the standard testing, which takes place in May, are usually reported back in December. As a result, in the vast majority of schools, namely in all primary and secondary schools (and partly also in academic secondary schools), the students of the tested classes have already left the school at the time of data feedback.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Herbert Altrichter http://orcid.org/0000-0002-5331-4199
Manuela Gamsjäger http://orcid.org/0000-0003-1174-3005
Julia Zuber http://orcid.org/0000-0003-2181-6968

References


Appendix 1. List of data

The quotation style for our data is as follows:

Case: PA = primary school A; PB = primary school B; PC = primary school C; SD = secondary school D; SE = secondary school E; AF = academic secondary school F.

Waves of data collection: 1 = 1st wave = 1st year; 2 = 2nd wave = 2nd year; 3 = 3rd wave = 3rd year.

Interviewees (identified by their functions):

- SL = School leader
- LI = Teacher
- LIT = Teacher teaching a class which participated in standard testing
- E = English teacher, if more than one is interviewed in a school: E1, E2, E3
- M = Math teacher
- D = German (i.e. language of instruction) teacher
- LD = “Learning designer”, i.e. teacher with special responsibility for classroom development (only in NMS)
- MF, DF, EF = subject coordinator (for Math, German, or English)
- SV = Student representative
- EV = Parent representative
- A few more special functions are not abbreviated
E.g., the code SE2/M2t. 4 refers to interview data which originates from secondary school E and from the second wave of data collection. The interviewee was a Math teacher who taught in a class that was subject to standard testing. Text in square brackets was inserted by the authors for explanation.

Note
German sources and data have been translated by the authors.