Unsatisfactory results in (international) student assessment served as a starting point for ongoing debates about the efficiency of the German school system. Subsequently, a new form of interplay between accountability and school autonomy has been advocated. Several instruments of a standards-based school reform have been implemented in Germany, such as statewide exams or school inspections. The data collected is not only used for monitoring by the administration, but should also inform decision making and practices in schools. In Germany, there is a discussion about whether and how data regarding the social environment of schools (e.g. economic and demographic data) or information concerning school or classroom composition should be used for resource allocation. At present, procedures for financing schools according to indicators and schools’ individual needs are exceptional cases in Germany (Morris-Lange, Wendt, & Wohlfarth, 2013). The federal states pursue different strategies; the city-state of Hamburg, for example, collects data systematically and uses this information for the allocation of resources. A purposeful allocation of resources is meant to do educational justice. However, there is little empirical evidence regarding the effectiveness of data-driven school improvement, and this is particularly true for the German context. Considering the expenses arising from extensive monitoring systems – i.e. public funds that could also be used for other purposes, such as the employment of additional teaching staff – proof of the usefulness of data is key. Moreover, research has focused mainly on the effects of individual instruments, such as school inspections or standardised testing, whereas comparative examinations of different sources of information have failed to appear. We try to close this research gap by investigating data use practices in German schools. What sources of information do practitioners consider useful, and which ones do they use? Furthermore, we are interested in the factors that (may) foster or hinder the use of evidence in schools. As the school context is supposed to influence data use, a comparison of schools in different settings seems to be a fruitful approach that has been largely neglected in previous research.

We draw on the findings of three different studies that were carried out in the German states of North Rhine-Westphalia and Rhineland-Palatinate, respectively. To gain a comprehensive picture of data use, these studies involved quantitative as well as qualitative data collection procedures. They examined the usefulness practitioners attach to particular types of evidence, as well as their actual usage of these information sources. We also examined reasons for and against the use of evidence in schools. To reveal a potential influence of the school context on
evidence-based school improvement, we compared data use practices in schools located in deprived areas to data use in schools in more favourable conditions. In the following article, we provide more information about the policy context in which our studies are situated. Therefore, we present a brief overview of previous research findings on data use. In the empirical part of the paper, we describe our conceptualisation of possible sources of evidence in schools. We outline the methods and data of our three studies, display the strategies for data analysis, and present our results. The paper closes with a summary and discussion of our findings.

**School autonomy and accountability in Germany**

As a consequence of sobering results in the Programme for International Student Assessment (PISA) 2000 study, extensive measures have been taken to improve the German educational system. Referring to Altrichter and Maag Merki (2010), three key measures can be distinguished: (1) increasing school autonomy, (2) applying managerial principles to schools, and (3) putting more focus on evidence-based decision making and evidence-based practice. According to the concept of evidence-based education, grounding actions and decisions in evidence are considered a prerequisite for efficient and effective practitioner performance, and an increase in student achievement (Honig & Coburn, 2008). Inspired by developments in the educational system of other countries, such as the United States and Great Britain, German policy makers have put a greater focus on the output and the outcome achieved. The corresponding change in the steering mechanism is reflected, for example, in the setting of educational standards and the implementation or reinforcement of standardised student testing or school inspections. The ‘pronouncement of new local autonomy and authority coupled with new centralised modes of evaluation (e.g. state inspectorates, central secondary school exit exams) and new technologies (e.g. national diagnostic tests, new evaluative authority over teachers)’ (Mintrop, 2015, p. 791) is also designed to contribute to principals and teachers behaving in a more managerial fashion.

The German Länder largely agree on these guiding principles to improve the educational system. Nevertheless, state-specific configurations can be observed. Rhineland-Palatinate, for instance, is the only state that has not implemented central exit exams at the end of secondary education. It also put an end to school inspections in 2016, so the stakes appear to be particularly low in this state. Compared to, for instance, the United States, accountability is low in Germany, as there are no penalties for low-performing schools. Schools are neither placed on probation nor closed following poor test results. Furthermore, insufficient results in student testing do not lead to the replacement of school leaders or teachers. A low degree of accountability also has consequences for data-driven school improvement. In this regard, practitioners’ willingness to use data seems to play an important role in trying to explain data use, as principals and teachers have room for manoeuvre. Moreover, although the individual teacher, principal, administrator or parent is the active party, accountability usually takes place at the school level (O’Day, 2004).

**Data use: a short review of the literature**

When it comes to research on data use in schools, inconsistencies in concepts and terms can be observed. In this regard, Coburn and Turner (2011, p. 175) described research on data use as ‘disorganized’; so far, a standard definition of evidence and evidence-based practice has not been developed. In a rather narrow sense, data can be understood as information about student or school achievement that has usually been gained by means of standardised testing. In a wider sense, evidence in schools comprises information such as social research findings, formative or summative evaluation data, school and student performance data, school statistics, school improvement plans, parent or student input, and comparing notes with peers (e.g. Honig & Coburn, 2008). Regarding research on the use of data in schools, three main research lines can be identified: (1) a comparably small body of research that investigates the relationship between initiatives to drive data use in schools and outcomes; (2) activities in the context of data use initiatives; and (3) normative pieces of writing that lack an analytical approach and are often very optimistic about data use and its effects (Coburn & Turner, 2012).

In Germany, most of the research on data use in schools has focused on the effects of single instruments of standards-based reform, such as national testing (e.g. Maier, 2009), exit exams (e.g. Kühn, 2010), or school inspections (e.g. Böttcher, 2007). Akin to the situation in the United States, other forms of data have been largely neglected (Bowers, 2009) and research findings comparing the use of different sources of information are extremely rare. However, in terms of encompassing school improvement, Schildkamp and Lai (2013) pointed to the necessity of using multiple sources of information:

> It is impossible to prescribe exactly which data schools must use, but (…) it is essential that school [sic] do not rely on one single source of
(achievement) data, but that they use multiple data sources: for example, the use of different sources of student (achievement) data and teacher (observation or survey) data (...). This is what researchers call triangulation. (p. 186)

The quote not only suggests looking at individual instruments for data-driven school improvement, but calls for a comparative examination of different sources of information. Creating an information-rich environment has also proven to be an effective strategy for the improvement of schools in challenging circumstances (Muijs, Harris, Chapman, Stoll, & Russ, 2004). For schools working in difficult conditions, Reynolds, Hopkins, Potter, and Chapman (2001) described a connection between data richness and student achievement. Thomas and colleagues (1998) also showed that schools that improved to a comparatively high degree were characterised by mobilising expertise for data use. Following the situational approach (also known as the contingency approach), student composition in schools in disadvantaged areas calls for differentiated practices (Reynolds, 1998). However, situational practices require knowledge about students’ resources, capabilities, performance, and progress. Here, the systematic utilisation of data can provide valuable clues.

A literature review reveals factors that (may) influence the use of data in schools. Here, three general factors can be distinguished (e.g. Coburn & Turner, 2011; Farley-Ripple & Cho, 2014):

(1) characteristics of the evidence itself,
(2) characteristics of the organisational context, and
(3) characteristics of the individual actor.

With respect to the characteristics of evidence, the availability, timeliness and validity of information sources has been determined. Evidence is used especially when practitioners perceive its relevance for their own work and conclusions are clear (Farley-Ripple & Cho, 2014). At the organisational level, school culture seems to play an important role in using evidence. Empirical findings have suggested that schools with a pronounced focus on innovation and flexibility use data to a greater extent. The same is true for schools with a very personal, family-like atmosphere (Demsiki, 2014). Furthermore, ‘technological, human, and resource capacity’ (Farley-Ripple & Cho, 2014) is conducive to the use of evidence, which suggests that the school context should be taken into account. Research findings have also pointed to the prominent position of school leaders in data-driven school improvement (e.g. Schildkamp & Lai, 2013; Wayman & Stringfield, 2006; Wohlsetetter, Datnow, & Park, 2008). They can provide resources (time, knowledge, space, or funds), and can enhance their staff’s use of evidence by raising expectations with respect to data-driven school improvement.

In this way, they can contribute to the development of professional learning communities and can foster collective sensemaking. These phenomena have proven to be influential in the use of evidence and school improvement in general (e.g. Schechter & Atarchi, 2014; Wayman, Jimerson, & Cho, 2012). As schools are situated in a multilevel system, evidence-based practice at the district level may also foster or hinder the use of evidence at the school level. For example, district leaders decide who gets access to which sources of information (Coburn & Turner, 2011). Therefore, the political context should also be considered when analysing data use in schools. At the individual level, the influence of beliefs, knowledge and individual experiences has been pointed out. Among others, the willingness to change as an integral part of practitioners’ professionalism, along with an advanced level of content knowledge (CK) and pedagogical content knowledge (PCK) for instructional development, provides supportive conditions for data use (Dederer, 2011). As Schildkamp and Lai (2013) noted, intrinsic rather than extrinsic motivation for data use is favourable: ‘The impetus for change should be driven from one’s own desire to use data, rather [than] driven by others’ (p. 182). Furthermore, the level of practitioner evaluation literacy seems to be significant when engaging with test data and research findings (e.g. Ikemoto & Marsh, 2007).

In Germany, studies investigating teachers’ use of data for school and classroom improvement show significant variation in the degree of reflection on data and the use of evidence-based information (Böttcher & Keune, 2012; Demsiki, in press; Kühl & van Ackeren, 2012). Although policy calls for the use of evidence in schools, empirical findings have indicated that data cannot be turned into successful evidence-based practice directly (Fleischman, 2009), as ‘people must actively make meaning of the data and construct implications for action’ (Coburn & Turner, 2011, p. 177). School improvement does not rely on educational research to the expected extent, which led Hargreaves (2007, p. 9) to the conclusion that it is ‘not good value for money’. Coburn and Turner (2011) noted that data – however called and however conceptualised – ‘are only as good as how they are used’ (p. 173). This means that analysing the extent to which data influence school leaders’ and teachers’ practices is crucial. Knapp, Copland, and Swinnerton (2007) described three tensions concerning data use, which – in their opinion – will always be present: (1) state/national policy vs. local response, (2) immediate feedback vs. longer-term documentation of performance, and (3) what is technically desirable vs. what is politically/culturally practicable. Analysing data use in schools can illustrate how the political claim for evidence-based education is put into practice.
Research questions

Little is currently known about what sources of information are used by practitioners, and this is especially true for the German context. Moreover, much of the recent research on data-driven decision making has focused on the effects of instruments of standards-based reform and, to our knowledge, comparative examinations of different sources of information are rare. We assume that the usefulness practitioners attach to an information source also affects its use. This leads to our first research question:

RQ 1: How much usefulness do practitioners attach to particular types of evidence?

Nonetheless, the provision of data may not be a sufficient precondition for evidence-based practice. Consequently, our second research question is:

RQ 2: What sources of information do teachers and school leaders use for their own work?

Furthermore, there is comparatively little evidence regarding factors influencing data-driven school improvement. This prompts the formulation of a third research question:

RQ 3: What reasons are put forward by school leaders and teachers for and against data use?

As there is limited knowledge about whether data use is dependent on the school context, we used a contrastive approach and compared our findings from schools in challenging circumstances to results from schools not located in deprived areas.

Data use in schools: operationalisation

A literature review pointed to data sources that may inform decision-making processes in schools. However, there is no standard definition in the literature, so we tried to conceptualise ‘data’ or ‘evidence’ as a first step. In order not to neglect possible ways of transferring credible findings about ‘what works’ into practice, we understand these terms in a rather broad sense. Table 1 provides an overview of different sources of information that might serve as evidence for practitioners (Demski, 2014). Here, we distinguish between evidence in a narrow sense and evidence in a broad sense. Evidence in a narrow sense is characterised by a relatively pronounced scientific focus and an external impetus, whereas evidence in a broad sense has a less pronounced scientific focus and is usually generated inside the individual school. Data can also focus on different levels, i.e. the system level, school/classroom level or teacher/student level. Moreover, an explicit or implicit potential to govern schools may underlie the various sources of information. Some procedures or information sources focus on processes in education (e.g. student feedback, collaborative measures), whereas others can be characterised as more output-oriented (e.g. statewide comparative tests or exit examinations).

Modes of inquiry and data

Frameworks on data use in schools and empirical findings have pointed out the influence of school culture and school context on evidence-based practice, which prompted us to compare schools in diverse environments. For this article, we draw on results from three different studies, which were carried out in the German states of North...
Rhine-Westphalia and Rhineland-Palatinate, respectively. North Rhine-Westphalia is the most heavily populated territorial state in Germany, whereas the population density is below average in Rhineland-Palatinate. Moreover, the unemployment rate is significantly higher in North Rhine-Westphalia (7.4% in October 2016) compared to Rhineland-Palatinate (4.8%; German average was 5.8% in October 2016).

Study 1, which was funded by the Federal Ministry of Education and Research, made use of a quantitative as well as a qualitative approach. We approached a random sample of 200 primary schools and invited all secondary schools in Rhineland-Palatinate to participate. Standardised questionnaires were applied to measure and describe the usage of available data for school improvement. A total of 1,230 teachers and 297 principals from 153 schools were surveyed about their attitude towards and their use of 13 information sources. The overall response rate in this study was 40.5% (teacher questionnaires) and 58.2% (school leader questionnaires), respectively. However, response rates varied between schools. Using five-point Likert scales, we examined the perceived usefulness and the self-reported use of the following sources of information in our questionnaires:

- international or national student assessments, e.g. PISA
- statewide comparative tests
- findings from school inspections
- school statistics
- collections of assignments/sample questions
- pedagogical journals related to a specific subject
- pedagogical journals not related to a specific subject
- articles in newspapers/magazines focusing on educational issues
- surveys carried out by the school/the teacher
- student feedback
- diagnostic tests in school/class
- collaborative measures, e.g. sitting in on each other’s classes
- school-based comparative tests

The interviewees came from six different types of schools. For the analyses presented in this paper we focused on secondary schools, excluding vocational schools and special needs schools. Thus, the sample was reduced to 424 teachers and 129 school leaders.

Moreover, 35 semi-structured, face-to-face interviews were conducted at seven schools (one school leader interview and four teacher interviews per school), to help further understand the causes and motivations for data-driven school improvement or neglect of existing information, respectively. The schools were selected using a contrastive approach and they differed in terms of the intensity of their evidence-based practice, as measured in the questionnaire study. There were four secondary schools employing intense use of evidence, and three secondary schools with a low use of evidence. The interviews were transcribed and analysed by means of qualitative content analysis (Mayring, 2007). The coding system was developed from the interview guidelines and was supplemented by categories that emerged when analysing the empirical data. We used software for qualitative data analysis (MAXQDA), and consolidated the interviewees’ arguments for and against data use into overarching categories.

Study 2 involved the evaluation of a programme that aims to sensitiise school leaders and teachers to issues of multiculturalism and multilingualism, and to enhance pupils’ competences in the language of instruction. Data were collected in all 33 North Rhine-Westphalian secondary schools participating voluntarily in the project, all of which were located in a socioeconomically deprived area³ and showed a high percentage of immigrant students and/or scored low on linguistic proficiency assessments. An online survey was conducted: 487 teachers and 46 principals assessed the usefulness, as well as the application, of the 13 different sources of information that were also of interest in the first study. The five-point Likert scales were also the same as in study 1. Due to the small school leader sample, we limited our analysis to teachers’ assessments. The overall response rate considering complete cases in the teacher questionnaire survey was 33%, with response rates ranging substantially, from 10% to 94%.

Study 3 was also funded by the Federal Ministry of Education and Research. Case studies were carried out in eight secondary schools, all of which were located in deprived areas in North Rhine-Westphalia. The schools were selected using a contrastive approach, considering the results achieved in statewide comparative tests in class 8 (so-called Vergleichsarbeiten, or VERA). Taking the classification by the Ministry for Schools and Further Education of the State of North Rhine-Westphalia into account, four of the selected schools performed as expected in the statewide tests. Another four schools were selected that performed unexpectedly well in the standardised testing of their students. The central agency in North Rhine-Westphalia that assures the quality of schools and offers support (called QUA-LiS NRW) was responsible for the selection of schools. In each school, one school leader interview and two teacher interviews were conducted. These interviews were transcribed and subjected to qualitative content analysis. We worked out the main reasons for and against the use of data in these interviews using MAXQDA. Here, we built on the coding system that was used for analysis of the face-to-face interviews conducted during study 1.
We used descriptive statistics to analyse the importance practitioners attached to the different types of evidence, and the degree to which they used the data. We then tested for statistically significant differences between schools in study 1 and the schools in disadvantaged areas in study 2. The main reasons for and against data use were identified by means of qualitative content analysis conducted in the course of study 1 and study 3. Here, we also looked for differences that could be traced to the school context.

Limitations

As there is no standard definition of evidence-based practice, our approach is only one of many possible procedures to measure this construct. Due to the social desirability response set, practitioners may have overrated their usage of evidence-based data and it may be asked whether a questionnaire survey is suitable to measure teachers’ and school leaders’ actual data use practice. In addition, it cannot be ruled out that a preponderance of evidence-based schools and practitioners participated in our surveys. Furthermore, the studies focused on the teacher level or school level, but in terms of educational governance (Altrichter, Brüsemeister, & Wissinger, 2007), other levels of the education system (e.g. the district level) may be influential as well. In addition, our comparison of disparate samples derived from various studies is just a first approach to dig into the influence of the school context on data-driven school improvement.

Results

The essential findings of the three studies are presented in the following section. First, we look at the perceived usefulness (RQ 1) and the actual use of the various sources of information (RQ 2). Here, we also examine whether there are any differences between schools in disadvantaged areas and schools in more favourable conditions. Finally, we present factors fostering or hindering evidence-based practice in schools as stated in the interviews (RQ 3). In this regard, we also compare our findings from study 1 to the results of study 3, which was conducted in schools in deprived areas.

RQ 1: how much usefulness do practitioners attach to particular types of evidence?

Results from the standardised questionnaires showed significant variance in the perceived usefulness and the use of evidence-based information (Figure 1). Moreover, high standard deviations indicate that the data points are spread out over a wider range of values. Usefulness was perceived to be highest for collaborative measures, student feedback and school-based comparative tests, whereas practitioners

![Figure 1. Usefulness of different information sources by school leaders (N_{study 1} = 129) and teachers (N_{study 1} = 424; N_{study 2} = 487), 1 (not useful at all) to 5 (very useful), mean values and standard deviations.](image)
attributed relatively little usefulness to the instruments of a standards-based reform. Helpfulness was perceived to be especially low by teachers for school inspection reports (2.18 on a five-point Likert scale in study 1; 2.91 in study 2). This was also true for school leaders’ assessments; findings from school inspections were rated least helpful by principals (2.72 on a five-point Likert scale). In general, teachers from schools in deprived areas attributed a higher degree of usefulness to the various information sources than did teachers working in schools located in more favourable conditions. Teachers and principals alike considered data that showed an immediate connection to teaching and learning particularly helpful. The strongest discrepancy between school leaders’ and teachers’ assessments in study 1 could be seen again in findings from school inspections and statewide, as well as school-based, comparative tests.

RQ 2: what sources of information do teachers and school leaders use for their own work?

Figure 2 shows the actual use of the different sources of information according to principals’ and teachers’ self-reports. Once again, the relatively high standard deviations are worth noting. Teachers and principals alike used students’ feedback and sitting in on each other’s classes frequently, whereas results from international or nationwide student assessments played only a minor role in school improvement. The latter was also true for findings from school inspections, so it can be concluded that these key policy instruments of standards-based reform are not used by practitioners to the desired extent. Teachers used subject-specific journals to a far greater extent than pedagogical journals not focusing on a specific subject; on the contrary, school leaders reported a comparable use of journals either relating or not relating to a specific subject. Apparent differences in principals’ and teachers’ assessments could be observed particularly with regard to findings from school inspections, school statistics and (inter)national student assessment. Thus, school leaders reported a considerably higher usage of standardised data that are rather output-oriented. Nevertheless, principals also reported basing their practice predominantly on information referring to processes in schools and showing an immediate connection to teaching and learning.

We were also interested in whether there were any differences in data use that may be conditioned by the school context. We tested for significant differences between schools in deprived areas (study 2) and schools not facing these challenging conditions (study 1) when it came to using the 13 information sources. We drew on teachers’ self-reports. Teachers from the

Figure 2. Use of different information sources by school leaders ($N_{study \ 1} = 129$) and teachers ($N_{study \ 1} = 424$; $N_{study \ 2} = 487$), 1 (not at all) to 5 (to a great extent), mean values and standard deviations.
former schools used results from statewide student assessment to a significantly ($p < .05$) higher degree. On the contrary, the use of diagnostic tests in school or in class, surveys carried out by the school or the teacher, collaborative measures, student feedback, pedagogical journals either related or not to a specific subject, articles in newspapers or magazines focusing on educational issues, school statistics, and collections of assignments were significantly higher in the schools in more advantageous conditions. No significant differences between the two groups could be found for the use of school inspection results, school-based comparative tests, and findings from international student assessment, such as PISA or Trends in International Mathematics and Science Study (TIMSS).

**RQ 3: what reasons are put forward by school leaders and teachers for and against data use?**

To improve understanding of the causes and motivations for data use, the 35 interviews conducted in the scope of study 1 and the 24 interviews from study 3 were analysed. We paid particular attention to factors fostering or hindering data-driven school improvement, and consolidated the interviewees’ arguments into overarching categories. The reasons for and against the use of evidence are presented in Table 2 and are illustrated in more detail thereafter.

**Factors fostering data use**

At the individual level, face-to-face interviews revealed different reasons for the use of data, with *ease of work* being mentioned most frequently. This was particularly true for the use of pedagogical journals focusing on a specific subject, as teachers were looking for practical teaching material. Nevertheless, it became obvious that deliberate use of journals to inform practice seems to be an exception:

> Well, if I read journals at all, then it’s not in a targeted way and I’m not looking for anything in particular. It’s as if I’m waiting at the dentist’s and flipping through the magazines and then something attracts my attention. I read it or I go on.

Only some teachers and school leaders stated the possibility of *scrutinising their own work* using evidence, so in these cases data use can be considered as a means of *professional development*. Whereas some teachers and school leaders mentioned a lack of trust among the staff and considered a competitive atmosphere in school to be an obstacle for collegial collaboration, other practitioners valued unbiased and objective evaluation by peers or school inspectors:

> (...) anyway, I think it’s good to scrutinise your own work from time to time so you don’t get stuck in your ways, and that you ask yourself every now and then, ‘Am I doing the right thing?’ or, ‘Could I try that out?’

School leaders in particular pointed out the possibility of using data for *signalling*, as outlined in principal–agent theory. Good results in statewide student assessments and school inspections were employed to demonstrate school quality and to attract high-performing students, as well as committed teachers. Practitioners in all schools taking part in the interviews conducted within the scope of study 1 stated increasing competition between schools, so this might become a more prominent argument in the future.

Some interviewees from schools in challenging circumstances mentioned the use of data concerning the social environment for purposes of resource allocation by educational administration authorities. These resources were used for targeted support of students in small learning groups, for example.

> Due to the high proportion of immigrant students, we get 40 extra school lessons in German that we can use for assistance in small groups.

Schools in deprived areas focused on data that provide information about their students’ performance level. The application of tests in particular for diagnosing students’ language level seems to be an established procedure.

> Schools in deprived areas focused on data that provide information about their students’ performance level. The application of tests in particular for diagnosing students’ language level seems to be an established procedure.

Well, I know that we use a language proficiency test in class 5. Based on the results, the students are grouped into German classes and German as a second language classes. And then a retest is conducted at the end of the school year, and based on this retest we ask ourselves which students still need this extra support in German.

In this case, the retest described goes beyond diagnosing learning conditions. It is used to check for the effectiveness of language development and to initiate further action. Long-term monitoring of cross-sectional student assessment data gives important evidence for the effectiveness of school improvement measures. To this end, schools use results from statewide comparative tests as seen, for example, in the following section from a teacher interview:

> And that is our explanation for the good results in the statewide comparative tests. (...) So, in our opinion, the reason for our improvement is to be seen
not only in more thorough diagnosis, but above all in the activities within these remedial lessons.

Besides results from statewide testing, school-based tests are used to compare the proficiency levels of different classes. These findings give cause to discuss striking differences between classes of the same year.

Factors hindering data use

A considerable proportion of the interviewees mentioned little data use due to a lack of time. Novice teachers in particular were busy with lesson planning and regarded themselves to be up to date in terms of pedagogical and content knowledge. Older teachers claimed to be experienced and mostly relied on their gut feeling rather than grounding their practice on available evidence. In the interviews, it became clear that practitioners meant different things when claiming to use student feedback. Moreover, it became clear that the use of student feedback was rarely institutionalised; rather, non-standardised procedures and methods were usually applied.

Furthermore, problems in recontextualising evidence and adapting it to one’s own needs were mentioned. Practitioners often did not realise the importance of data and feedback for their own work. Poor test results were usually not related to teachers’ own practice, but attributed to students’ poor cognitive skills. It also became obvious that practitioners usually did not adapt their teaching practices to their students’ needs and requirements.

(... that I could use it for my teaching specifically, I don’t know right now, I can’t see an immediate connection.

Moreover, some interviewees stated shortcomings in the quality and validity of the data, which was particularly true for findings from school inspections and standardised testing of students. Practitioners expressed a wish for disaggregated data on the student level and custom-fit evaluation and feedback. Instead of summative performance reviews, process support and consulting were favoured:

Anyway, I don’t consider this snapshot very fruitful. That would require a process, an ongoing study or something, but it doesn’t work by observing lessons for 20 minutes every now and then and interviewing the teaching staff every other year, that’s my personal view.

Some practitioners from schools in deprived areas criticised instruments for standardised student testing due to floor effects. They noted a lack of opportunities for differentiation at the bottom end of the performance range, partly due to the considerable proportion of students not achieving even the lowest competence level in these standardised tests. In this case, feedback to students and parents is hardly helpful as it focuses exclusively on deficits:

I have to provide opportunities for success to students; there is no use in solely highlighting shortcomings.

When analysing the face-to-face interviews, the influence of leadership and school culture on data-driven school improvement also became obvious. Schools showing intense usage of evidence had created a cooperative, family-like atmosphere in which the principal functioned as a mentor who promoted data-driven school improvement. Theses principals were often data-wise themselves; nevertheless, they pursued a rather distributed leadership style and delegated responsibilities to working groups (Demski & Racherbäumer, 2015). In contrast to the less evidence-based schools, the schools showing a comparatively high degree of data use seemed to value the benefits of data, rather than seeing them as an additional workload. They did not consider testing as a means of monitoring by the school supervising authority, but a possibility to foster organisational learning and to demonstrate school quality in terms of signalling.

Discussion

Knowledge about what sources of information are of greatest use for practitioners is rather limited, and this is particularly true for the German context. This article has tried to decrease this research gap by providing empirical findings about data use in German schools. In general, principals used data focusing on the school level more frequently than did teachers, possibly because they mediate between administration and teaching staff. Nevertheless, the analyses showed that internally gathered sources of information – data that we called evidence in a broad sense – were especially useful for teachers and principals. External data were considered less useful and consequently played only a minor role in school improvement. Thus, information about the output achieved does not automatically inform the processes that preceded it, and professional practice. As stated in the interviews, internal instruments and data met the requirements of the individual school more effectively than did highly standardised feedback instruments. This leads to the question of how effective the instruments of a standards-based reform actually are. The interviews revealed that most practitioners had difficulties in recontextualising external data to utilise it for their own practice; they also questioned the usefulness of these data. Hence, these findings not only pose the question of how useful standardised data are for practitioners, but also raise the issue of capacity-building for data
use. Our analyses showed that teachers in study 2, who were working in schools in deprived areas, tended to report a higher level of perceived usefulness for the various sources of information compared to their colleagues in study 1. However, teachers’ data use in study 2 was less pronounced, so data use cannot be regarded exclusively as a matter of attitude. Highly aggregated data can provide valuable knowledge for educational administration and educational policy. Nevertheless, these data may be less influential for school leaders’ and teachers’ practice and decision making because, according to practitioners, they lack cultural relevance for the specific context of an individual school (Heinrich, 2013). Taking up the conceptualisation and understanding of evidence as a knowledge base that is subject to communication, sensemaking and recontextualisation, it can be assumed that these highly standardised and aggregated data do not resonate well with schools in challenging circumstances. If practitioners working in schools in deprived areas get the impression that their specific situation is ignored in data collection and feedback, they may prove defiant. Indeed, we were able to observe negative attitudes and defensive reactions in our studies. Therefore, taking practitioners’ points of view into account and understanding evidence as a result of communication and negotiation is a promising approach that could open up new avenues for data-driven school improvement (Bremm, Eiden, Neumann, Webs, & van Ackeren, 2017). Against this backdrop, the way in which data feedback could be modified in order to give rise to discussion, negotiation and collective sensemaking in schools must be examined. This could result in eventual school improvement. In this regard, academic support seems necessary to drive data use in schools; moreover, further research is needed to disclose what actually happens when practitioners engage with data.

Educational policy turns out to be particularly effective when considering practitioners’ interests and logic (Fend, 2008). Therefore, ‘data-based evidence should not only be descriptive, but should generate and provide knowledge [on] how to initiate change’ (Demski, 2014, p. 145). The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (Kultusministerkonferenz, KMK) has reacted to this and improved its strategy for educational monitoring (KMK, 2015). Now, a greater focus is placed on linking findings to explanations and recommendations for action. In this respect, researchers in education should also scrutinise the usefulness of their results for practitioners. Usually, research findings are not linked to recommendations for action and researchers mainly publish their work for the scientific community, not for practitioners. Furthermore, the wealth of research findings and the fact that data analysis in educational research is becoming increasingly complex (e.g. multilevel or structural equation modelling) complicates practitioners’ ability to notice and reflect on data. To counteract this possibility, systematic reviews and research syntheses may be helpful. In addition, the development of subject-specific materials seems to be an approach worth discussing, as it could enhance the relevance of evidence that teachers perceive. Our analyses also showed that journals covering a specific subject were used to a far greater extent than journals not relating to a subject. Currently, the Technical University of Munich, for instance, aims to establish a clearing house in which evidence about effective teaching and learning in STEM subjects is edited and disseminated.³

A comparison of the different samples indicated that data use in general might be lower in schools in challenging circumstances. Therefore, implementing programmes to enhance data use in these schools might be worth discussing; however, far more research is needed to analyse evidence-based practice in schools, especially in the case of disadvantaged areas. Our findings point to the need for systematic and custom-fit support structures that also take the school context into account. By building practitioners’ capacity for internal assessment, teachers might also become less timid of engaging with standardised test data. Moreover, the possible benefits of evaluation and feedback should be made clearer to schools. Even though the German educational system can be considered a low-stakes environment, the interviewees in our surveys using evidence to a minor degree assessed the control function of standardised tests and school inspections more highly than the support function. Thus, it may be questioned whether stronger coupling between accountability and evaluation necessarily results in evidence-based school improvement. Moreover, the risk of ‘collateral damage’ (Nichols & Berliner, 2007) has been illustrated in this regard. The fear of consequences following poor results might be particularly pronounced in schools facing challenging circumstances. Apart from that, the interviews displayed different strategies for a systematic monitoring of student progress that are used to foster the development of teaching. Moreover, the floor effects mentioned in the course of statewide comparative tests should be analysed further. Altogether, the interview studies illustrate that schools in deprived areas predominantly collected and used student performance data. On a positive note, this result shows that schools strive to collect and use student performance data systematically, and there is therefore a foundation for data use to build on. On the downside, it
becomes obvious that schools are very focused on student achievement, neglecting possible interrelationships between teaching practice and student performance. Especially for schools in challenging circumstances, other sources of information, such as data related to well-being in schools or learning strategies, could be starting points for school improvement as well. As Schildkamp and Lai (2013) pointed out, using multiple data sources is essential for analysing the actual situation and for improving the quality of schools. Findings from the interview studies also indicate that practitioners tend to attribute bad test results to students’ poor cognitive skills; poor results do not usually cause teachers to challenge their own work. This approach is particularly opposed to the improvement of schools in deprived areas, and reveals the need for a change in practitioners’ beliefs and attitudes that should already be conveyed in teacher training. Moreover, more strongly embedding the principles of evidence-based school improvement in teacher training could result in greater openness towards using evidence, and greater evaluation literacy.

In this context, another problem with limiting the use of evidence to student performance data becomes obvious. As Altrichter (2009) stated, the assumption that the standardised testing of students leads to an improvement in professional practice is based upon a rather unusual understanding of feedback. The fact that teachers do not receive immediate feedback about their work, but should derive implications for actions from feedback concerning their students’ performance, can explain why practitioners in our studies had problems recontextualising evidence, and why they favoured information sources such as student feedback or collaborative measures.

A further problem in the context of data use might derive from the fact that setting standards and imposing data use in top-down processes goes together with little intrinsic motivation in most cases. Other authors have also pointed out this issue:

A problem in English schools is that the goals of data use, according to most school staff, are primarily external (data use for accountability). However, effective data use, in our opinion, can only be achieved if the goals are primarily internal: for example, data use to increase student achievement on mathematics with at least 5% in one school year. (Schildkamp & Lai, 2013, p. 182)

Specifying organisational targets and linking these targets to monitoring and data use might enhance practitioners’ commitment to data use. This might also be more encouraging than the negative feedback concerning student achievement and process quality with which schools in challenging circumstances are often confronted. Furthermore, our studies indicate that leadership and school culture might influence evidence-based practice in German schools. Therefore, school improvement should also consider school culture and possibilities for developing culture. We have shed light on these correlations in more detail elsewhere (Demski, 2014, in press).

Still, there are desiderata to be answered in future research. Our data is based on practitioners’ self-reports, so data use might be overrated in our studies. Coburn and Turner (2012, p. 99) stated: ‘In particular, we still have shockingly little research on what happens when individuals interact with data in their workplace settings.’ This calls for qualitative approaches to disclosing individual and collective sensemaking processes (for instance, through observations of formal and informal settings). In this regard, the potential benefit of micro-process studies (Little, 2012) can be pointed out, which might also help to disclose the influence of the school context on data use. Furthermore, proof of a positive impact of evidence-based school development on student achievement is still lacking.

Notes
1. Figures retrieved from www.statistikportal.de on 29 November 2016.
2. For our assessment of the school context, we drew on the classification by the Ministry for Schools and Further Education of the State of North Rhine-Westphalia, which classifies secondary schools in North Rhine-Westphalia into five groups according to official statistics QUA-LiS NRW (2017). The variables that are taken into account include the proportion of immigrant students among pupils, as well as the unemployment rate and the proportion of people receiving social welfare in the school environment. This classification also serves as a basis for fair comparisons of schools concerning results in statewide tests.
3. See https://www.edu.tum.de/qualitaetsoffensive/clearing-house-unterricht/ for details (retrieved 27 November 2016).

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by the Bundesministerium für Bildung und Forschung [01JG1008; Bundesministerium für Bildung und Forschung [01JG1010B]; Stiftung Mercator [12-348-1].

ORCID
Denise Demski http://orcid.org/0000-0002-9795-9837
References

Altrichter, H. (2009). Datenfeedback und Unterrichtsentwicklung. Probleme eines Kernelements im ‘neuen Steuerungsmodell’ für das Schulwesen [Data feedback and improvement of teaching. Problems of a core element in the new steering mechanism of the educational system]. In W. Böttcher, J. N. Dicke, & H. Ziegler (Eds.), Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspraxis und pädagogischer Praxis (pp. 211–226). Münster: Waxmann.

Altrichter, H., Brüsmeister, T., & Wissinger, J. (Eds.). (2007). Educational Governance. Handlungskoordination und Steuerung im Bildungssystem [Educational Governance. Coordination of action and steering in the educational system]. Wiesbaden: VS.

Altrichter, H., & Maag Merki, K. (2010). Steuerung der Entwicklung des Schulwesens [Steering of the development of the school system]. In H. Altrichter & K. Maag Merki (Eds.), Handbuch Neue Steuerung im Schulsystem (pp. 15–39). Wiesbaden: VS.

Böttcher, W. (Ed.). (2007). Schulinspektion: Evaluation, Rechenschaftslegung und Qualitätsentwicklung [School inspection: Evaluation, accountability, and quality]. Münster: Waxmann.

Böttcher, W., & Keune, M. (2012). Externe Evaluation und die Steuerung der Einzelschule: Kontrolle oder Entwicklung? [External evaluation and steering of schools: Control or development?]. In S. Ratermann & S. Stöbe-Blossey (Eds.), Governance von Schul- und Lernumfelden (pp. 63–80). Wiesbaden: Springer VS.

Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. Journal of Educational Administration, 47(5), 609–629. doi:10.1108/09578230910981808

Bremm, N., Eiden, S., Neumann, C., Webs, T., & van Ackeren, I. (2017). Evidenzbasierter Schulentwicklungsansatz für Schulen in herausfordernden Lagen. Zum Potenzial der Integration von praxisbezogener Forschung und Entwicklung am Beispiel des Projekts ‘Potenziale entwickeln – Schulen Stärken’ [Evidence-based school improvement for schools in challenging circumstances. On the potential of integrating practice-oriented research and development using the example of the project Developing potentials – Strengthening schools’]. In P. Dobbelstein & V. Manitzus (Eds.), Schulentwicklungsarbeit in herausfordernden Lagen (pp. 140–158). Münster: Waxmann.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. Measurement: Interdisciplinary Research and Perspectives, 9(4), 173–206.

Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction. American Journal of Education, 118(2), 99–111. doi:10.1086/663272

Dedering, K. (2011). Hat Feedback eine positive Wirkung? Zur Verarbeitung extern erhobener Leistungsdaten in Schulen [Does feedback have a positive effect? Processing of external assessment data in schools]. Unterrichtswissenschaft, 39(1), 61–81.

Demski, D. (2014). Which data do principals and teachers use to inform their practice? Evidence from Germany with a focus on the influence of school culture. In A. J. Bowers, A. R. Shooh, & B. G. Barnett (Eds.), Using data in schools to inform leadership and decision making (pp. 121–150). Charlotte, NC: Information Age.

Demski, D. (in press). Evidenzbasierte Schulentwicklung – Empirische Analyse eines Steuerungsparadigmas [Evidence-based school improvement – Empirical testing of a ‘steering paradigm’]. Wiesbaden: Springer VS.

Demski, D., & Racherbäumer, K. (2015). Principals’ evidence-based practice – Findings from German schools. International Journal of Educational Management, 29(6), 735–748.

Farley-Ripple, E. N., & Cho, V. (2014). Depth of use: How district decision-makers did and did not engage with evidence. In A. J. Bowers, A. R. Shooh, & B. G. Barnett (Eds.), Using data in schools to inform leadership and decision making (pp. 229–252). Charlotte, NC: Information Age.

Fend, H. (2008). Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität [Shaping schools. System control, school improvement, and instructional quality]. Wiesbaden: VS.

Fleischman, S. (2009). User-driven research in education: A key element promoting evidence-based education. In W. Böttcher, J. N. Dicke, & H. Ziegler (Eds.), Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis (pp. 69–82). Münster: Waxmann.

Hargreaves, D. H. (2007). Teaching as a research-based profession: Possibilities and prospects. In M. Hammersley (Ed.), Educational research and evidence-based practice (pp. 3–17). London: Sage.

Heinrich, M. (2013). Bildungsgerechtigkeit für alle! aber nicht für jeden? Zum ‘Individual-Disparitäten-Effekt’ als Validitätsproblem einer Evidenzbasierung [Educational justice for all – but not for everyone? On the ‘individual-disparities-effect’ as an issue of validity of evidence-based practice and decision making]. In F. Dietrich, M. Heinrich, & N. Thieme (Eds.), Bildungsgerechtigkeit jenseits von Chancengleichheit. Theoretische und empirische Erginungen und Alternativen zu ‘PISA’ (pp. 181–194). Wiesbaden: Springer VS.

Honig, M. I., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Towards a policy and research agenda. Educational Policy, 22(4), 578–608. doi:10.1177/0895904807307067

Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the ‘data-driven’ mantra. Different conceptions of data-driven decision making. In P. A. Moss (Ed.), Evidence and decision making. The 106th yearbook of the National Society for the Study of Education (pp. 105–131). Malden, MA: Blackwell.

KMK (Ed.). (2015). Gesamtrategie der Kultusministerkonферenz zum Bildungmonitoring [Overall strategy for educational monitoring of The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany]. Retrieved August 12, 2015, from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtratenge-Bildungsmonitoring.pdf

Knapp, M. S., Copland, M. A., & Swinnerton, J. A. (2007). Understanding the promise and dynamics of data-informed leadership. In P. A. Moss (Ed.), Evidence and decision making, 106th yearbook of the national society for the study of education (pp. 74–104). Malden, MA: Blackwell.

Kühle, B., & van Ackeren, I. (2012). Wirkungen externer Evaluationsformen für eine evidenzbasierte Schul- und Unterrichtsentwicklung [Effects of external evaluation for evidence-based school improvement]. In M.
Governance von Redesigning accountability systems for School Effectiveness and School Improvement (pp. 45–62). Wiesbaden: VS. Kühn, S. M. (2010). Steuerung und Innovation durch Abschlussprüfungen? [Governance and innovation by means of exit exams?]. Wiesbaden: VS.

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. American Journal of Education, 118(2), 143–166. The practice of data use. doi: 10.1086/663271

Maier, U. (2009). Wie gehen Lehrerinnen und Lehrer mit Vergleichsarbeiten um? Eine Studie zu testbasierten Schulexperimenten in Baden-Württemberg und Thüringen [How do teachers deal with comparative tests? A study in the course of test-based school reforms in Baden-Wuerttemberg and Thuringia]. Balthmannsweller: Schneider.

Mayring, P. (2007). Qualitative Inhaltsanalyse. Grundlagen und Techniken [Qualitative content analysis. Foundations and techniques] (9th ed.). Weinheim: Deutscher Studien Verlag.

Mintrop, R. (2015). Public management reform without managers: The case of German public schools. International Journal of Educational Management, 29 (6), 790–795.

Morris-Lange, S., Wendt, H., & Wohlfarth, C. (2013). Segregation an deutschen Schulen. Ausmaß, Folgen und Handlungsempfehlungen für bessere Bildungschancen [Segregation in German schools: extent, consequences, and recommendations for action]. The Expert Council of German Foundations on Integration and Migration (Ed.), Berlin. Retrieved October 12, 2016, from http://www.svr-migration.de/wp-content/uploads/2013/07/SVR-FB_Studie-Bildungssegregation_web.pdf

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas? A review of research evidence. School Effectiveness and School Improvement, 15(2), 149–175. doi:10.1076/sesi.15.2.149.30433

Nichols, S. L., & Berliner, D. C. (2007). Collateral damage. How high-stakes testing corrupts America’s schools. Cambridge, MA: Harvard Education Press.

O’Day, J. (2004). Complexity, accountability, and school improvement. In S. H. Fuhrman & R. F. Elmore (Eds.), Redesigning accountability systems for education (pp. 15–43). New York, NY: Teachers College press.

QUA-LiS NRW. (2017). Deskriptive Beschreibung der Standorttypen für die weiterführenden Schulen [Description of location types of secondary schools]. Retrieved August 23, 2017, from http://www.schulentwicklung.nrw.de/e/upload/lernstand8/download/mat_2017/2017-02-08_Beschreibung_Standorttypen_wiederhrende_Schulen_NEU_RUB_ang.pdf

Reynolds, D. (1998). The study and remediation of ineffective schools: Some further reflections. In L. Stoll & K. Myers (Eds.), No quick fixes: Perspectives on schools in difficulties (pp. 163–174). London: Falmer Press.

Reynolds, D., Hopkins, D., Potter, D., & Chapman, C. (2001). School improvement for schools facing challenging circumstances: A review of research and practice. London: Department for Education and Skills.

Schechter, C., & Atarchi, L. (2014). The meaning and measure of organizational learning mechanisms in secondary schools. Educational Administration Quarterly, 50(4), 577–609. doi:10.1177/0013161X13508772

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. K. Lai, & L. M. Earl (Eds.), Data-based decision making in education. Challenges and opportunities (pp. 177–191). Dodrecht: Springer.

Thomas, G., Davies, J., Lee, J., Postlethwaite, K., Tarr, J., Yee, W. C., & Lowe, P. (1998). Best practice amongst special schools on special measures: The role of action planning in helping special schools improve. Bristol: University of Bristol.

Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the Data-Informed District. School Effectiveness and School Improvement, 23(2), 159–178. doi:10.1080/09243453.2011.652124

Wayman, J. C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. American Journal of Education, 112(4), 549–571. doi:10.1086/505059

Wohlsetetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. School Effectiveness and School Improvement, 19(3), 239–259. doi:10.1080/09243450802246376