Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 8 (2016), No. 3, 14–38

Christian Dormann, Carmen Binnewies, Anna Rosa Koch,
Isabell van Ackeren, Marten Clausen, Peter Preisendörfer,
Uwe Schmidt & Olga Zlatkin-Troitschanskaia

Transferring best evidence into practice: Assessment of evidence-based school management

Abstract

Improving the quality of the German educational system has become a pivotal interest among several stakeholders. In parallel, the concept of evidence-based management (EBMgt) has emerged. The present paper deals with establishing the concept of evidence-based school management (EBSMgt). Previous research has established EBMgt as consisting of three dimensions: (a) external evidence orientation (EE), (b) internal evidence orientation (IE), and (c) evidence substitute orientation (ES; Stumm, Mohr, & Dormann, 2010). We applied the scales developed by Stumm et al. to a sample of N = 2,573 teachers and to N = 296 school principals and their deputies employed at N = 168 schools in the state Rhineland-

Prof. Dr. Christian Dormann (corresponding author) · Prof. Dr. Olga Zlatkin-Troitschanskaia, Faculty of Law, Management and Economics, Chair of Business Education and Management, University of Mainz, Jakob-Welder-Weg 9, 55128 Mainz, Germany
e-mail: cdormann@uni-mainz.de
troitschanskaia@uni-mainz.de

Prof. Dr. Carmen Binnewies · Anna Rosa Koch, Institute of Psychology, Work Psychology, University of Muenster, Fliednerstr. 21, 48149 Muenster, Germany
e-mail: carmen.binnewies@uni-muenster.de
anna.r.koch@uni-muenster.de

Prof. Dr. Isabell van Ackeren, Faculty of Educational Sciences, Chair of School Development Research, University of Duisburg-Essen, Campus Essen/Weststadttürme, 45117 Essen, Germany
e-mail: isabell.van-ackeren@uni-due.de

Prof. Dr. Marten Clausen, Educational Research Workgroup, Department for Instructional Research, University of Duisburg-Essen, Campus Essen/Weststadttürme, 45117 Essen, Germany
e-mail: marten.clausen@uni-due.de

Prof. Dr. Peter Preisendörfer, Faculty Social Sciences, Media, and Sports, Institute of Sociology, University of Mainz, Jakob-Welder-Weg 12, 55128 Mainz, Germany
e-mail: preisendoerfer@uni-mainz.de

Dr. Uwe Schmidt, Center for Quality Assurance and Development, University of Mainz, Colonel-Kleinmann-Weg 2, 55128 Mainz, Germany
e-mail: Uwe.Schmidt@zq.uni-mainz.de

Palatinate (RLP) of Germany. The types of schools considered reflect 92.3 % of the schools in RLP. Principal component analyses for teachers and for principals confirmed the three-dimensional structure of EBSMgt. Alphas were satisfactory for IE (teachers: .87; principals: .74) and for EE (.79 & .69, resp.), but they were lower for ES (.53 & .59, resp.). Furthermore, variance component analyses revealed shared perceptions within schools (ICC1) for EE. ICC1 for IE and ES was weaker. Multilevel modeling revealed meaningful relations for the three scales with variables used for validation purposes (e.g., previously used evidence and usefulness ratings of evidence). Dichotomizing the three scales and cross-tabulation yielded several EBSMgt "types", of which the ideal EBSMgt type (high IE & EE, low ES; 19.75 %) and the evidence-averted type (low IE & EE, high ES; 18.47 %) were most common. A major contribution of the present study is the provision of the three scales, which allow assessing EBSMgt in terms of different types of school management. This provides a foundation for future studies to identify ways to improve EBSMgt and to investigate its various consequences.

Keywords
Evidence-based management; Schools; Best practice; Local evidence

Transfer wissenschaftlicher Evidenz in die Praxis: Messung evidenzbasierten Schulmanagements

Zusammenfassung
Die Verbesserung der Qualität des deutschen Schulwesens hat eine hohe Priorität für diverse Interessensvertreter erlangt. Gleichzeitig entwickelte sich das Konzept des evidenzbasierten Managements (EBMgt). Der vorliegende Aufsatz zielt darauf ab, das Konzept des evidenzbasierten Schulmanagements (EBSMgt) zu etablieren. Vorhergehende Forschung hat gezeigt, dass EBMgt aus drei Dimensionen besteht: (a) externe Evidenzorientierung (EE), (b) interne Evidenzorientierung (IE) und Orientierung an Evidenzsubstituten (ES; Stumm, Mohr & Dormann, 2010). Wir haben die Skalen von Stumm et al. in einer Stichprobe von N = 2573 Lehrern/innen und N = 296 Schulleitern/innen und ihren Stellvertretern/innen in N = 168 Schulen in Rheinland-Pfalz (RLP) validiert. Die berücksichtigten Schultypen repräsentierten 92.3 % der in RLP vorkommenden Schulen. Hauptkomponentenanalysen der Daten von Lehrern/innen und Schulleitern/innen bestätigten die dreifaktorielle Struktur. Für IE (Lehrer/innen: .87; Schulleiter/innen: .74) und EE (.79 bzw. .69) waren die Alphas zufriedenstellend, wohingegen sie für ES geringer waren (.53 bzw. .59). Weiterhin zeigten Varianzkomponentenanalysen, dass innerhalb von Schulen bezüglich EE geteilte Wahrnehmungen existierten (ICC1), wohingegen die ICC1 bei IE und ES geringer ausfiel. Multilevelanalysen ergaben erwartungskonforme Beziehungen der drei Skalen mit Variablen, die zur Kriteriumsvalidierung eingesetzt wurden (z. B. bereits zuvor genutzte Evidenz oder Nützlichkeitseinschätzungen von Evidenz).

Dichotomisierung und Kreuztabellierung ergaben verschiedene EBSMgt-„Typen", von denen der Idealtyp (hohe IE und EE, geringe ES; 19.75 %) und der evidenzabgewandte Typ (geringe IE und EE, hohe ES; 18.47 %) am häufigsten vorkamen. Ein wesentlicher Beitrag zur Forschung liegt darin, dass mit den drei Skalen ein Instrument bereitgestellt wird, um EBSMgt als unterschiedliche Typen des Schulmanagements zu erfassen. Damit ist eine Grundlage dafür geschaffen, in weiteren Studien Möglichkeiten zur Verbesserung von EBSMgt zu identifizieren und verschiedenartige Konsequenzen von EBSMgt zu untersuchen.

Schlagworte
Evidenzbasierte Steuerung von Schulen; Best Practice; Lokale Evidenz

1. Introduction

Over the last decades, ample empirical evidence on best practices to improve school effectiveness has been gathered and is constantly refined by pilot studies as well as small-scale and large-scale scientific investigations. The transfer of available evidence into everyday practice, however, rarely is satisfactory and frequently causes a huge knowledge-doing gap (Pfeffer & Sutton, 2000). To overcome this gap, evidence-based management (EBMgt) has been suggested as a means for organizations to succeed (Pfeffer & Sutton, 2006). EBMgt means focusing on existing empirical evidence and "translating principles based on best evidence into organizational practices" (Rousseau, 2006, p. 256). However, evidence-based school management (EBSMgt), that is, schools and school principals acting in a way that the transfer of best evidence into their schools' practices is supported, has not yet received much attention. The present paper aims at filling in this gap by developing and validating the concept of EBSMgt and its measurement.

Evidence-based development of schools is a hot topic among many stakeholders involving, for example, researchers, politicians, local authorities, teachers' seminaries, and schools (van Ackeren et al., 2011). Although a huge body of evidence regarding effective practices does already exit, several studies report a limited use of existing evidence by teachers and school principals (e.g., Posch, 2009). Reasons for this limited use are multifold and have been enumerated by management and educational scholars (e.g., Briner, 2007; Maier & Kuper, 2012). For instance, access to well translated evidence has been frequently claimed as an *exogenous* reason, that is, a reason outside of organizations and schools. However, even if well translated evidence is available, using evidence is sometimes inconvenient or even pretty demanding as it frequently implies that the status quo has to be changed (Briner, 2007), Resistance to change (e.g., Battistelli, Montani, & Odoardi, 2013) of those who are affected can be assumed to be a strong *endogenous* counter-force to evidence-based management. Overcoming resistance to change requires excellent leadership (Bass & Riggio, 2006), and we propose that evidence-oriented lead-

ership is required to acquire, use, and implement evidence in an organization, be it either profit-organizations or schools. Thus, we propose EBSMgt to go above and beyond particular strategies and use of particular data to improve school development. The concept of EBSMgt is much broader and comprises of general patterns of individual and collective ways of data use and decision-making in schools. EBSMgt represents a general managerial style that sustains all aspects of school development.

1.1 Evidence-based school management (EBSMgt)

Some research results suggest that the management style in schools may either encourage or discourage the use of evidence (cf. Wayman, 2005). Wayman (2005) proposed that a "leadership for supportive data climates" (p. 302) leads to a school culture that will increasingly involve teachers in making use of evidence. To achieve this, principals should serve as role models and they should support and encourage teachers in using evidence. However, to our knowledge there has been no attempt yet to empirically establish EBSMgt as a managerial style that sustains school development. Contrary to this, since the turn of the century we recognize an immense growth of interest in the concept of EBMgt. Reay, Berta, and Kazman Kohn (2009) identified 144 papers dealing with EBMgt since 1948, and roughly 90 % were published after the turn of the century. The concept of EBMgt has at latest become popular by the work of Pfeffer and Sutton (2006). EBMgt represents a normative leadership theory (Externbrink & Dormann, 2015) stating that decisions taken by leaders of organizations, institutions, branches, schools etc. should be founded on the best scientific evidence that is available (Rousseau, 2006). Many scholars have claimed that leaders frequently ignore existing evidence when judging alternative options for actions. Rather, leaders tend to base their decisions on so-called evidence substitutes. Evidence substitutes include, for example, gut feelings, intuitions, blind benchmarking, personal opinions, and repeating what has been done in the past (Pfeffer & Sutton, 2006).

Distinguishing evidence from evidence substitutes is not as trivial as it might appear at a first glance. Early literature has mainly regarded scientific evidence, that is, research results, as "real" evidence. More recently, Briner, Denyer, and Rousseau (2009) proposed three additional evidence sources beyond *research results* that should be considered and integrated in organizational decision-making: Practitioner *expertise and judgment*, the *local context*, and *stakeholders' feedback*. Therefore, school principals' expertise, the actual procedures and processes in a given school, and students' feedback could be regarded as evidence, too.

Regarding school principals' expertise and judgment or students' feedback as similar good evidence as results from scientific studies might be a provocative claim for some scholars. However, regarding only scientifically generated knowledge as evidence is provocative for practitioners, too. Practitioners are convinced

that their decision-making is based on evidence, and claims to be more evidence-oriented are meaningless to them (Briner & Rousseau, 2011). The key question is: When and why should practitioner expertise and stakeholders' feedback be regarded as "valid evidence"?

Externbrink and Dormann (2015; see also van Ackeren et al., 2013) claimed that it does not matter whether the evidence emerges from scientific research, practitioners' expertise, or stakeholders' feedback. Rather, they suggested regarding any sort of information as evidence if it is objective, reliable, and valid. Information is objective if independent recipients agree, information is reliable if it can be replicated, and information is valid if it informs the recipients about what they really want to know. It would lack a scientific argument to regard results of a survey of pupils as evidence if the survey was carried out by researchers, whereas the same kind of study would be regarded as evidence substitutes if school teachers designed the survey and analyzed the data. Provided the teachers were educated in social science methods, the survey could be equally objective, reliable, and valid. Validity, of course, is the key issue. For instance, students' satisfaction could be objectively and reliably assessed, but it could be entirely invalid to inform decisions which teaching method is efficient.

Practitioners' expertise and stakeholder feedback may even be more valid than results from academic research projects. As Briner et al. (2009) emphasized, local context plays an important role, and Pfeffer and Sutton (2006) already noted that blind benchmarking, that is, just doing what successful others do without considering possible differences in contexts, represents an evidence substitute. Blindly using findings from academic research for school development purposes is conceptually close to blind benchmarking. Instead, the local context of a particular school needs to be carefully considered. If the context of an academic study does not resemble the context of a particular school, even not fully valid principals' expertise or not fully valid students' feedback could be more useful than only internally valid but externally invalid (i.e., not generalizable) academic research results.

Consideration of local expertise and judgment, and stakeholders' feedback is also considered to be of pivotal importance in contemporary pedagogy (e.g., Gruber, Harteis, & Rehrl, 2006) and is referred to as the ideographic approach. According to the ideographic approach, pedagogical professionalism is characterized by a comprehensive analysis of the specific pedagogical case at hand. The purpose is not to derive general propositions, which would resemble the common nomothetic approach as used by many researchers. An ideographic analysis should include all relevant case-specific information. This implies that teachers' expertise with a particular case, for example, a student's learning history, her family issues, her social relations in class, her own aims and desires, etc. is taken into account when making decisions, such as whether she should repeat a grade or not. An ideographic analysis, may well come to conclusions that divert from scientific evidence. Hence, professional pedagogical actions cannot solely be based on quantitative research results; rather professional acting requires expertise and experience-based reflection (van Ackeren et al., 2011).

1.2 Assessment of EBSMgt

Although the literature on EBMgt and EBSMgt has been exploding in recent years, there still is little evidence on EBMgt (Externbrink & Dormann, 2015; Reay et al., 2009). One core problem for this unfortunate state of affairs is the lack of assessment tools and measures of EBMgt and EBSMgt. Obviously, it is a complex endeavor to make an overall evaluation of the *evidence grade* of a given school, or to assess the extent to which all recently made decisions and measures were evidence-based or not.

In order to assess the extent to which local authorities act evidence-oriented, Stumm, Mohr, and Dormann (2010) developed a measurement tool. Employees and managers were asked to rate a range of activities related to evidence-based actions and decisions. Several of those activities indicate that evidence is important in their area (e.g., "In our administration, innovations are proved by scientific studies."), others indicated that using evidence substitutes is common (e.g., "The statement 'We always did it this way' is the basis for many decisions."), and there were also items addressing consideration of the local context (e.g., "Before adopting procedures from other administrations, we analyze if our framework conditions are similar."). Stumm et al. (2010) indeed identified a three factor-solution. In their terms, EBMgt consists of three components: high *external evidence orientation* (EE; research evidence, contact with scholars, etc.), high *internal evidence orientation* (IE; considering local context, own research, etc.), and low *evidence substitute orientation* (ES; gut feeling, tradition, etc.).

The four building blocks of EBMgt, suggested by Briner et al. (2009; research results, expertise and judgment, local context, and stakeholders' feedback), are represented across the two factors EE and IE identified by Stumm et al. (2010). EE consists of research results and scientists' expertise and judgment; IE consists of employees' and managers' expertise and judgment, local context, and systematically evaluated stakeholders' feedback. Stumm et al. (2010) also noted that the presence of EE-based or IE-based decisions does not imply that evidence substitutes are neglected; rather ES represents a factor that is conceptually independent of IE and EE. IE (ICC1 = .14) and EE (ICC1 = .12) showed sufficient interrater reliability (James, 1982) among employees, whereas interrater reliability was low for ES (Stumm et al., 2010).

1.3 Research aims

We based our study on the work of Stumm et al. (2010). Their newly developed scales represent a sound basis to assess EBMgt in schools because they are concise and coherently cover the full range of EBMgt proposed by Briner et al. (2009). Thus, the aim of our study was to adapt the measures of Stumm et al. (2010) to the school context and to empirically address the question of whether EBSMgt can

be reliably and validly assessed by teachers and by principals. Note that the items developed by Stumm et al. (2010) were different for employees and their managers. Whereas the managers were asked to rate the way *they* decide and act (i.e., the items usually contained an "I"), employees were asked to rate the way *their managers* decide and act (the items usually contained "our manager/s", "in our administration", or "here"). We kept these features in our questionnaires because we aimed at assessing EMSMgt at the school level and not at the individual level of teachers' behavior. Thus, a first set of variables addressed IE, EE, and ES of *principals* and deputies, and the extent of *teachers' perceptions of principals* in being IE, EE, and ES.

Secondly, for validation purposes, we were interested in a couple of very particular sources of evidence (e.g., nationwide learning assessments) that have *actually been used* in previous school development activities. For further validation purposes, we analyzed relations with the *perceived usefulness* of these particular sources of evidences. To test our different hypotheses, the three sets of variables or combinations of them were either analyzed on the *individual level* of teachers or at the *aggregated level* of schools.

Thirdly, we were interested in how the evidence factors that we could empirically identify, combine, and yield different "types" of EBSMgt. In particular, we were interested in the frequencies of prototypical types of EBSMgt, such as types characterized by high evidence orientation and low substitute orientation.

Our first two hypotheses state that the three factor structure discovered by Stumm et al. (2010) can also be replicated in schools.

- Hypothesis 1: Teachers' perceptions of principals' decision-making and ways of acting in their schools can adequately be represented by three factors: IE, EE and ES.
- Hypothesis 2: Principals' ratings of their own decision-making and their own way of acting can adequately be represented by three factors: IE, EE and ES.

In order to further validate the EBSMgt concept and its respective scales, we aimed at establishing a nomological network. Therefore, we analyzed correlations with possible meaningful antecedents and consequences of evidence orientation and substitute orientation.

Participation in previous evidence-oriented school development activities represents such a meaningful antecedent variable. In particular, having previously *used* evidence-oriented information such as those obtained in school assessments should increase teachers' perceptions that their schools are more evidence-oriented and less substitute-oriented. Specifically, we examined the use of information from *external* sources (e.g., statewide learning assessments) and from *internal* sources (e.g., student surveys) as antecedents of EBSMgt.

To establish convergent and discriminant validity, we propose that previous use of external evidence impacts more strongly on teachers perceptions of principals' EE; previous use of internal evidence should impact more strongly on teachers' perceptions of principals' IE. Furthermore, the use of information from both exter-

nal and internal sources should reduce perceptions that the own school is substitute-oriented.

- Hypothesis 3: Teachers' use of internal evidence information from previous school development activities increases teachers' perceptions that their principals exhibit more IE (convergent validity). This relation is stronger than the relation between use of internal evidence information and EE (discriminant validity) and stronger than the relation between use of internal evidence information and ES (discriminant validity).
- Hypothesis 4: Teachers' use of external evidence information from previous school development activities increases teachers' perceptions that their principals exhibit more EE (convergent validity). This relation is stronger than the relation between use of external evidence information and IE (discriminant validity) and stronger than the relation between use of external evidence information and ES (discriminant validity).

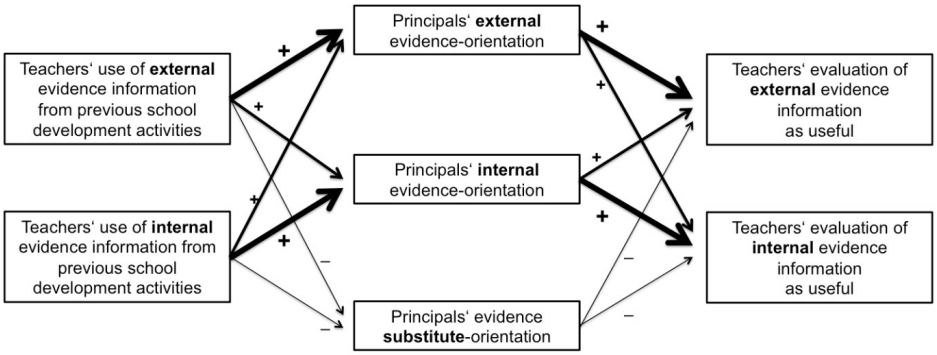
Furthermore, principals and other members of the school management team act as role models for teachers. The more teachers perceive their school leaders to be evidence-oriented and not to be substitute-oriented, the better should teachers judge the *usefulness* of evidence information. Hence, whereas previous *use* of evidence should increase teachers' perceptions of their principals' evidence orientation, perceiving their principals as evidence-oriented should impact on teachers' ratings of the *usefulness* of evidence. We hypothesize:

- Hypothesis 5: Teachers perceiving more IE among their principals evaluate internal evidence as useful (convergent validity). This relation is stronger than the relation between IE and the usefulness of external evidence, and it is stronger than the relation between IE and the usefulness of evidence substitutes (discriminant validity).
- Hypothesis 6: Teachers perceiving more EE among their principals evaluate external evidence as useful (convergent validity). This relation is stronger than the relation between EE and the usefulness of internal evidence, and it is stronger than the relation between EE and the usefulness of evidence substitutes (discriminant validity).

Our hypotheses are summarized in Figure 1. Differences in the widths of the paths indicate expected differences in absolute size of the expected regression coefficient.

Thus far, we have used terms such as EBSMgt and evidence *orientation*. Based on the work by Stumm et al. (2010), van Ackeren et al. (2013) proposed that different combinations of these orientations yield different types of evidence-based *management*. They distinguished four types of information seeking behavior in schools, focusing more or less strongly on evidence and on evidence substitutes. They termed the management style that is evidence-oriented and at the same time not substitute-oriented as EBSMgt. Indeed, this combination reflects the ideal prototype of EBSMgt. Van Ackeren et al. (2013) also noted that being interested in evidence substitutes in addition to evidence is not necessarily a bad thing because evidence substitutes are not *per se* incorrect information; rather, it is not known if they are correct or not. Thus, schools high in both evidence orientation and evi-

Figure 1: Research model and hypotheses



dence substitute orientation are schools that are open to any kind of information that could be used for school development purposes. Such a type of school management was termed *volatile school management*. The opposite style was termed *sclerotic school management*. Sclerotic school management rarely uses any information, be it evidence or evidence substitute-information. Such schools might have little interest in any sort of information that might be useful for school management, or they might be schools that indeed engage in few school development activities. Finally, the management style of schools that make use of evidence substitutes, but fail to consider evidence, was labeled *evidence-averted school management*. Although this style may seem to be irrational at first glance, it may be the result of few available resources, high daily demands, convenience, etc.

In our paper, we aim at determining the types of school management represented in our sample. Above and beyond the work by van Ackeren et al. (2013), we are firstly interested in considering all factors or dimensions that represent EBSMgt in our sample. Secondly, we also aim at a simultaneous consideration of teachers' and principals' perceptions. Thirdly, we do not want to set arbitrary cut-off values for low vs. high evidence orientation but prefer to make an empirically based distinction (i.e., using a median split). This led to Research Question 1: How frequent are different types of EBSMgt.

2. Method

2.1 Sample

We collected data in N = 168 German schools located in the state of Rhineland-Palatinate. Study participants were German school principals and teachers from different school types. We limited the sample to the major school types in the state, which represented 92.3 % of all schools. We approached all schools by sending

study information and directly speaking to principals. At the school level, 18.3 % participating schools were primary schools (out of 56.4 % at the state level), 23.5 % were schools for students with special needs (out of 8.3 %), 17.6 % were junior high schools (out of 9.5 %), 12.4 % were high schools (out of 8.8 %), 7.2 % were integrated schools (out of 3.1 %), and 20.9 % were vocational schools (out of 6.1 %).

Within the participating schools, the overall response rate was 41 %. Principals and their deputies showed a higher response rate (57 %) than teachers (40 %). Sample sizes differed, however, among different types of analyses. The items used to measure EBSMgt (see below) were applied in all questionnaires. For analyses involving only EBSMgt-items, up to N = 2,573 teachers and up to N = 296 school principals provided data. For further validation analyses, sample size was smaller because some items were only applied in a randomly selected_subsample involving about 50 % of participants.

In the overall sample, most participants worked at primary schools (*Grundschulen*, 27.9 %) and vocational schools (*Berufsschulen*, 24.3 %), followed by junior high schools (*Realschulen* & *Realschulen Plus*, 15.3 %), high schools (*Gymnasien*, 13.4 %), schools for students with special needs (*Förderschulen*, 11.1 %), and integrated schools (*Gesamtschulen*, 5.8 %). Among the participating teachers, 61.9 % were female. Among principals and their deputies, 39.5 % were female. Average age was 44.28 years (SD = 10.45) for teachers and 51.77 (SD = 8.01) for principals and deputies. Teachers worked at the current school for an average of 12 years (SD = 10). Principals worked at their current school for an average of 13 years (SD = 9) and were employed as deputy or principal for an average of 9 years (SD = 5).

2.2 Measures

2.2.1 Items to assess EBSMgt

Items used to assess EBSMgt were adopted from Stumm et al. (2010). The wording of all items answered by teachers is shown in the Results section in Table 1. For principals and their deputies, words such as “we” or “in our school” were replaced by “I”. Responses were made on a scale from 1 (*I do not agree at all*) to 5 (*I totally agree*). Psychometric properties are provided after the principal component analyses in the Results section.

2.2.2 Variables used for validation of EBSMgt (nomological network)

Used evidence: We proposed that teachers who previously used evidence to improve their own work have higher perceptions of the level of evidence orientation in their schools. In particular, we were interested in two external sources:

Nationwide learning assessments (*used external assessments*) and school inspections and external evaluations (*used external inspections*); two other were internal sources: School-level parallel exams within grades (*used internal exams*) and students' feedback on lessons (*used internal feedback*). We did not include variables that addressed previous use of information from substitute-like school-development activities because we felt this could not be properly done in a questionnaire. The verbatim wording in the questionnaire for teachers was "Have the following procedures/tools been available or have they been used?" Responses required marking of either *no* (0) or *yes* (5). For used external assessments, ICC1 = .27, for used external inspections, ICC1 = .50, for used internal exams, ICC1 = .37, and for used internal feedback, ICC1 = .22.

Evidence usefulness: We proposed that teachers having higher perceptions of the level of evidence orientation in their schools regard information from preceding school development projects as more useful. As before, we were interested in teachers' usefulness ratings of two external sources: nationwide learning assessments (*external assessments usefulness*) and school inspections and external evaluations (*external inspections usefulness*); again, two other were internal sources: school-level parallel exams within grades (*internal exams usefulness*) and students' feedback on lessons (*internal feedback usefulness*). The verbatim wording in the questionnaire for teachers was "How do you rate the in general the usefulness of the following procedures/tools for working at a school?" Responses were made on a 5-point scale ranging from 1 (*not useful at all*) to 5 (*very useful*). For usefulness of external assessments, ICC1 = .07, for usefulness of external inspections, ICC1 = .12, for usefulness of internal exams, ICC1 = .15, and for usefulness of internal feedback, ICC1 = .02.

3. Results

To establish the factorial structure of evidence orientation in schools, in a first step we analyzed the principal components of the data obtained from teachers. One item had around 7 % missing values (166) and we decided to exclude it from further analyses (ES10: "In our school, guest observations in other schools with similar problems are regarded as more valuable than further professional education."). The scree plot showed a clear bend between the 3rd and the 4th principal component, so we decided to limit the extraction to three factors. The factor solution for the teacher data (N = 2,021) is shown in Table 1.

Analyzing the data obtained from school principals, a first principal component analysis yielded five factors with eigenvalues > 1.0. There was not a clear bend in the scree plot, but the eigenvalue of the third factor (1.84) was clearly higher than from the fourth factor (1.39). Furthermore, since a three factor solution promised the best comparison between teachers and principals, we decided to extract three

factors. The results of the exploratory factor analysis for the school principal data (N = 262) is also shown in Table 1.

The factor solutions for teachers and principals were quite similar. Remember that the items answered by teachers referred to how decisions are made by their principals (or "in our school"), whereas the items answered by the principals referred to themselves ("me" or "I").

The first factor was marked by items referring to the creation, use, and critical evaluation of *internal evidence*. There were three small differences for the pattern of loadings between teachers and principals. Item EOH1 (The principal's personal beliefs are the foundation of the decision) had similar loadings on the first and the second factor for teachers. For principals, these items loaded mainly on the second factor. Since the fit in terms of content is closer to the remaining items of the second factor (to be discussed below), we decided to exclude EOH1 from the subsequently computed scale based on the first factor. The third item, EOH14, had a single loading above .40 only for teachers but not for principals. Furthermore, for teachers this item had a cross-loading on the second factor and it had the lowest loading of the remaining items marking on the first factor. Finally, the content of this item was not as clear as were the remaining ones. Therefore, EOH14 was also excluded from computation of the final scale, which was labeled *internal evidence orientation* (IE), and which comprised items EOH11, EOH15, EOH18, EOH27, EOH29, ES5, and ES6.

The second factor was again comparable for teachers and principals and described the reliance on *evidence substitutes*. The items loading on this component reflect instinct and gut feeling. For teachers, items ES3, EOH1, and EOH14 had cross-loadings on the first factor. Item EOH14 (doing what other schools do) did not have a major loading on this factor for principals and was thus discarded. Although item ES3 (decisions are based on personal experiences) had a cross-loading on the first factor for teachers. It matches the content of this factor and was thus retained. It is interesting to note that three items referred to personal experiences, of which one marked the factor IE (EOH11: "When it comes to important decisions, figures based on experience are of great relevance in our school.") and the other two loaded on ES (ES3: "The majority of decisions made in our school are based on personal experiences." & ES11: "Based on broad experience, in our school correct decisions are made intuitively."). A possible explanation is that ES3 and ES11 regard experience more as the unquestioned mode of any kind of decision-making, which reflects what Salas, Rosen, and DiazGranados (2010) termed *immature intuition*. On the other hand, in EOH11 experience reflects perhaps more a kind of *expertise-based intuition* (Salas et al., 2010) that comes into play when very important decisions are made. Item EOH1 (decisions are based on personal beliefs of the principal) was similar to ES3, and since it also had a strong loading on this factor for principals we decided to keep ES3 for the subsequently computed second scale. Thus, finally, we used items ES3, ES4, ES8, ES9, ES11, and EOH1 to subsequently compute the second scale, which was labeled *evidence substitute orientation* (ES).

Table 1: Factor solution of oblique principal component analysis of teacher data (N = 2,021) and principal data (N = 262)

		Teachers			*Principals*		
Label	Content	F1	F2	F3	F1	F2	F3
EOH2	Before important innovations are introduced, we explicitly search for verified (e.g., scientific) evidence of their efficacy in our school.			**.63**			**.63**
EOH6	Our school principals have direct contact to researchers to improve the quality of their decisions.			**.78**			**.80**
EOH10	When it comes to important decisions the expertise of consultants is requested in our school.			**.72**			**.65**
EOH11	When it comes to important decisions, figures based on experience are of great relevance in our school.	**.68**			**.52**		
EOH15	By recognizing inconvenient truths and facts we can learn a lot about errors and their prevention.	**.46**			**.51**		
EOH18	Before our school implements new methods and rules, we analyze their efficacy.	**.61**			**.54**		
EOH27	In our school information is retrieved from various sources before processes are re-designed.	**.57**			**.51**		
EOH29	Before adopting procedures from other schools, we analyze if our framework conditions are similar.	**.72**			**.68**		
EOH30	In our school, innovations are proved by scientific studies.			**.74**			**.69**
EOH31	Our school conducts development projects and research projects together with university students and doctoral students.			**.57**			**.69**
ES5	Before adopting procedures from other schools, we ask ourselves why it was successful there.	**.66**			**.69**		
ES6	Before we introduce new methods, we try to imagine possible shortcomings, even if we favor the idea.	**.76**			**.67**		
ES3	The majority of decisions made in our school are based on personal experiences.	-.45	**.57**			**.55**	
ES4	In our school we trust the gut feeling when it comes to important decisions.		**.69**			**.60**	
ES8	The statement 'We always did it this way' is the basis for many decisions in our school.		**.51**			**.43**	
ES9	In our school the content of advanced training frequently is less important than exchanging experiences with colleagues.		**.50**			**.51**	
ES11	Based on broad experiences, in our school correct decisions are made intuitively.		**.47**			**.74**	
EOH1	Decisions are based on the personal beliefs of our school principals.	-.48	**.47**			**.65**	
EOH14[a]	Decisions in our school are made based on what is done in other schools.	-.45	.42				

Note. Loadings < .40 in absolute value are not shown for the ease of interpretation. The matrix of full loadings is available from the first author. Loadings printed in bold face indicate the final assignment of items to scales (see text for further explanation).
[a] This item was finally excluded from the computation of scales.

The third factor, which comprised items referring to the use of external evidence (EOH2, EOH6, EOH10, EOH30, and EOH31), was again comparable for teachers and principals. The content of these items refers to evidence created outside the or-

ganization and made available, for example, by exchange and cooperative research projects with researchers and consultants. We subsequently computed a scale labelled *external evidence orientation* (EE), which comprised items EOH2, EOH6, EOH10, EOH30, and EOH31.

In the next step, we analyzed the reliability of the three scales for teachers. Alpha for the scale IE was .87 (ICC1 = .09), alpha for EE was .82 (ICC1 = .12), and alpha for ES was .57 (ICC1 = .05).

The reliabilities and ICCs of the three scales for principals were as follows: Alpha for IE was .74 (ICC1 = .00), alpha for EE was .73 (ICC1 = .20), and alpha for ES was .64 (ICC1 = .04).

To summarize the results obtained thus far, we have found clear evidence for the proposed three-factorial structure of EBSMgt among teachers, which confirms Hypothesis 1. The number of factors for principals was not as clear, however, the three-factorial structure obtained was very close to the structure obtained using teachers' ratings. Thus, Hypothesis 2 was at least partially supported.

In the next step, we computed the descriptive statistics (means, standard deviations, and correlations) of the new scales measuring the orientation towards IE, EE, and ES in the management of the schools from the principals' and the teachers' perspectives. First, we determined the correlations based on *individual* perceptions, which yielded the correlations among the three scales for teachers and among the three scales for principals, respectively. These are shown in the top panel of Table 2. Then we *aggregated* teachers' and principals' data at the school level and computed the correlations based on aggregated data. This yielded the correlations among the three scales for teachers and principals. They are shown in the bottom panel of Table 2.

At the individual level the correlations between the IE and EE on the one hand and ES on the other hand were negative as expected. Among teachers, the three correlations were higher in absolute value than among principals. Again, correlations between IE and EE on the one hand and ES on the other hand were negative.

For school level data, a similar pattern of correlations within principals and within teachers was observed. The correlations between IE and EE on the one hand and ES on the other hand were again negative as expected. In general, the correlations among the principals' variables, which ranged from $r = .09$ to $r = .35$ in absolute value, were generally lower than the correlations among the teachers' variables, which ranged from $r = .31$ to $r = .60$ in absolute value. Nevertheless, correlations were not extraordinarily high so that treating the scales as separate constructs is well-justified.

The convergent validities, that is, the pairwise correlations of the three principals' scales with the three teachers' scales were positive as expected. For IE, however, correlations between aggregated teachers' and principals' scores were not significant. Noteworthy, principals' self-reported use of EE was negatively ($r = -.16$, $p < .05$) correlated with teachers' reports of their principals' ES.

Table 2: Means, standard deviations, and correlations of main study variables

No.	Scale	*M*	*SD*	1	2	3	4	5	6
						individual level			
1	internal evidence (P)	4.15	.47						
2	external evidence (P)	2.95	.69	.37**					
3	substitute orientation (P)	2.89	.48	-.10+	-.17**				
4	internal evidence (T)	3.44	.72						
5	external evidence (T)	2.60	.78				.61**		
6	substitute orientation (T)	2.86	.53				-.27**	-.33**	
						school level (aggregated data)			
1	internal evidence (P)	4.13	.35						
2	external evidence (P)	2.94	.55	.35**					
3	substitute orientation (P)	2.89	.36	-.09	-.19*				
4	internal evidence (T)	3.52	.32	.06	-.03	-.03			
5	external evidence (T)	2.86	.40	-.02	.19*	-.10	.60**		
6	substitute orientation (T)	2.64	.25	-.10	-.16+	.17*	-.32**	-.31**	

Note. N = 291–292 for principals in the top panel. N = 2,450–2,483 for teachers in the top panel. N = 148–154 for schools in the bottom panel.
** $p < .01$, * $p < .05$, + $p < .10$.

The reliabilities and the pattern of correlations provided further partial support for Hypotheses 1 and 2. The ICCs were not always as high as desirable – we will discuss this in the Discussion section.

According to Hypothesis 3 and Hypothesis 4, previous use of information that emerged from evidence-oriented measures such as school assessments should increase teachers' perceptions of their principal's evidence orientation and decrease their perceptions of principal's substitute orientation. Furthermore, principals' act as role models, and Hypothesis 5 and Hypothesis 6 proposed that teachers who perceive their principals to be evidence-oriented and not to be substitute-oriented would regard evidence more positive than teachers whose principals are less evidence-oriented and more substitute-oriented. To test these hypotheses, we used the ratings provided by teachers and analyzed them at the individual level. The correlations among the three EBSMgt scales, *used evidence,* and *evidence usefulness* are shown in Table 3. All correlations were positive and mostly significant except those correlations involving evidence-substitution, which were negative and one half of them significant.

To simultaneously test Hypothesis 3 to Hypothesis 6, we further analyzed a two-level multiple mediation model using structural equations. We used MPLUS and simultaneously considered within and between school relations (cf. Zhang, Zyphur, & Preacher, 2009). To estimate the relations within schools, variables were centered at the respective school mean (group mean centering); to estimate relations between schools, grand mean centering was used. Previous use of evidence

Table 3: Means, standard deviations, and correlations of teachers' EBM-scales, previous use of evidence, and perceived usefulness of evidence (individual level, teacher data)

No	Scale	*M*	*SD*	1	3	2	4	5	6	7	8	9	10
1	used ext. assessments	1.98	1.05										
2	used ext. inspection	2.89	1.37	20**									
3	used int. exams	2.77	1.36	43**	15**								
4	used int. feedback	3.53	1.15	14**	13**	12**							
5	ext. assessments usefulness	2.94	1.21	27**	08*	07*	02						
6	ext. inspection usefulness	2.42	1.19	22**	23**	05	06	39**					
7	int. exams usefulness	3.79	1.15	22**	09**	50**	07*	42**	21**				
8	int. feedback usefulness	4.14	.88	05	08*	08**	54**	06	12**	15**			
9	int. evidence orientation	3.44	.69	14**	06*	17**	15**	11*	21**	24**	16**		
10	ext. evidence orientation	2.62	.76	19**	04	14**	13**	07*	26**	15**	09**	61**	
11	substitute orientation	2.85	.53	-06	-04	-10**	-05	-03	-17**	-12**	-04	-32**	-38**

Notes. Listwise $N = 974$ teachers. Decimals in the correlation table were omitted.
** $p < .01$, * $p < .05$.

information served as the independent variable, teachers' perceptions of their principals' IE, EE, and ES as mediators, and perceived usefulness of evidence information as the dependent variable.

First, we report results obtained from the *within-schools* sub model. Table 4 shows the results from multiple mediation analysis. The grey-shaded cells of Table 4 display the convergent validities. Convergent validity was established for both scales measuring the use of internal evidence, which impacted significantly on IE but failed to significantly impact on EE and ES. Thus, Hypothesis 3 was clearly supported. For the use of external evidence results were mixed. With respect to external inspections, that pattern of coefficients was as expected, albeit the difference between EE and IE was small. However, concerning the use of external assessments, results did not confirm expectations. Thus, overall, Hypothesis 4 was only partially supported.

Hypothesis 5 stated that teachers perceiving their principals as more internally evidence-oriented evaluate internal evidence as useful. This relation should be stronger than the relation between IE and the usefulness of external evidence, and it should also be stronger than the relation between IE and the usefulness of evidence substitutes. Hypothesis 5 was partly supported. The effect of IE on the use-

Table 4: Within-school and between-school path coefficients in a two-level mediation model (Teacher Data)

Sources	Outcomes unstandardized coefficients *(SE)* *t*-value						
	internal evidence orientation (IE)	external evidence orientation (EE)	evidence substitute orientation (ES)	usefulness internal parallel exams	usefulness internal feedback	usefulness external assessments	usefulness external inspections
				within schools			
used internal parallel exams	.38 (.07) 5.76**	.02 (.06) .31	-.07 (.06) -1.08				
used internal feedback	.30 (.05) 5.63**	-.01 (.05) -.32	.05 (.05) 1.06				
used external assessments	.25 (.07) 3.80**	.07 (.07) 1.08	.12 (.08) 1.47				
used external inspections	.21 (.06) 3.27**	.24 (.06) 4.28**	-.14 (.07) -2.12*				
internal evidence orientation (IE)				.05 (.02) 2.97**	.08 (.02) 4.06**	-.01 (.03) -.26	.09 (.02) 3.70**
external evidence orientation (EE)				.03 (.02) 2.04*	.06 (.02) 2.84**	.04 (.03) 1.29	.12 (.03) 4.47**
evidence substitute orientation (ES)				-.04 (.02) -2.62**	-.00 (.02) -.12	.02 (.02) .59	-.06 (.02) 3.10**
				between schools			
used internal exams	.34 (.24) 1.42	-.34 (.23) -1.44	-.46 (.33) -1.42				
used internal feedback	.10 (.14) .76	.03 (.14) .18	.01 (.16) .06				
used external assessments	-.02 (.22) -.09	-.21 (.24) -.88	-.25 (.24) -1.01				
used external inspections	.23 (.56) .90	.01 (.31) .04	-.51 (.29) 1.77				
internal evidence orientation (IE)				-.01 (.04) -.16	.16 (.08) 2.00*	.12 (.09) 1.34	12. (.06) 2.09*
external evidence orientation (EE)				-.01 (.05) -.09	.14 (.09) 1.52	.09 (.10) .82	.15 (.07) 2.23*
evidence substitute orientation (ES)				-.03 (.03) -1.05	.01 (.05) .22	.04 (.06) .67	-.08 (.04) -2.14*

** $p < .01$, * $p < .05$.

fulness of both sorts of internal evidence was significant as expected. However, the effect on the usefulness of external inspections was stronger and significant, too. Similarly, Hypothesis 6 was partly supported, too. The effect of EE on the usefulness of external inspections was significant and the strongest of the four effects. However, the effect on the usefulness of external assessments failed to reach significance, whereas the effects on the usefulness of both sorts of internal information

were significant. We had no specific hypotheses regarding the effects of teachers perceiving their principals as evidence substitute-oriented. However, not unexpectedly, teachers perceiving their principals as evidence substitute-oriented regard external inspections and internal parallel exams as less useful.

The bottom part of Table 4 shows the results from the *between-schools* sub model. In brief, there was neither support for Hypothesis 3 nor for Hypothesis 4, and there was partial support for Hypothesis 5 and Hypothesis 6. Results also show that in schools where teachers perceive their principals as evidence substitute-oriented, external inspections are evaluated as less useful.

Finally, we addressed Research Question 1 to determine what types of school management exist and how frequent they are. We considered all three dimensions that we discovered before. To simultaneous consider teachers' and principals' perceptions, we firstly aggregated the data at the school level, and then we computed the mean of the aggregated teachers' and principals' scores. Then we determined the median school level for each of the three scales, and we used dichotomization to define if a school is either high or low in any of the three orientations.

Table 5: Types of evidence-oriented and evidence substitute-oriented schools

		low evidence substitute orientation		
		internal evidence orientation (IE)		
		low	high	total
external evidence orientation (EE)	low	25 (31.60 %) *15.92 %*	13 (16.50 %) *8.28 %*	38 (48.10 %) *24.20 %*
	high	10 (12.70 %) *6.37 %*	31 (39.20 %) *19.75 %*	41 (51.90 %) *26.12 %*
	total	35 (44.30 %) *22.29 %*	44 (55.70 %) *28.03 %*	79 (100 %) *50.32 %*
		high evidence substitute orientation		
		internal evidence orientation (IE)		
		low	high	total
external evidence orientation (EE)	low	29 (37.20 %) *18.47 %*	17 (21.80 %) *10.83 %*	46 (59.00 %) *29.30 %*
	high	17 (21.80 %) *10.83 %*	15 (19.20 %) *9.55 %*	32 (41.00 %) *20.38 %*
	total	46 (59.00 %) *29.30 %*	32 (41.00 %) *20.38 %*	78 (100 %) *49.68 %*

Note. Percentages in parentheses indicate percentages within the respective panel (low vs. high substitute orientation), percentages in italic indicate overall percentages.

Cross-tabulation yielded the results shown in Table 5. The top part of Table 5 shows the results for schools that were characterized by low substitute orientation. The most frequent type was the type which reflects prototypical evidence-based school management, and it was found in nearly 40 % of the schools, followed by the sclerotic type (31.60 %). Types that exhibited a mixture of high and low EE and IE were less frequent (12.70 % and 16.50 %).

The bottom part of Table 5 shows the results for schools that were characterized by high substitute orientation. The most frequent type was the evidence-averted type found in 37.20 % of the schools. The relative frequencies of the other three types were lower and quite similar (in between 19.20 % and 21.80 %).

4. Discussion

The aim of the present study was to further conceptually and empirically investigate the construct of EBSMgt. The concept of EBSMgt goes above and beyond particular strategies and use of particular data to improve things at school; rather EBSMgt represents a general managerial style that sustains all aspects of school development. The EBSMgt concept is founded on the concept of EBMgt and its measurement instrument developed by Stumm et al. (2010). The present study demonstrated that EBMgt and EBSMgt share conceptual and empirical characteristics. Principal component analyses revealed a three factor-structure, which was very similar for teachers and for principals. In line with Stumm et al. (2010), who found three factors among employees in public administrations, we identified three factors, too. These three factors were IE, EE, and ES.

Internal consistencies were generally acceptable except for the ES scales (.57 for teachers and .64 for principals). Thus, to better assess the substitute orientation in school management, the ES-scales warrant further improvement. It could be, however, that improving the reliability turns out to be difficult. Just adding further items might not yield the expected result because there might not be a clear criterion for demarcation according to which principals' decisions can be clearly defined as being based on evidence or on evidence substitutes. For instance, in a recent case, a German school principal raised the high-school (Abitur) grades for exams in German by 1 point for all students (Lehmann, 2013). Is this measure evidence-based or based on gut feeling? Presumably, many readers do not believe it was evidence-based. However, the principal used to be a teacher of German language, so he certainly has had some solid expertise. Representatives of the responsible ministry of education have confirmed that the marks awarded by the first two teachers who rated the written exams of the students were indeed too low (BR, 2013). Thus, in general, observers may or may not know if a decision is based on immature intuition that is better regarded as evidence substitute, or based on expertise-based intuition (Salas et al., 2010), which could reflect the *best available evidence*. Sometimes there is a narrow ridge between evidence and evidence substitutes, so

that reliably assessing whether or not decisions were made based on substitutes could be rather difficult.

Intra-class correlations for EE were satisfactory for teachers (.12) and principals (.20), for IE it was marginally satisfactory for teachers (.09) but not for principals (.00). For ES, ICCs were generally low (.05 for teachers and .04 for principals). Despite the low ICCs for ES, aggregated school level data show that teachers' and principals' perceptions are correlated (r = .17). This is also the case for EE (r = .19). Overall, convergent validities were not very high, which was not unexpected because low correlations between self-reports and other reports are a well-established fact in the literature (e.g., on 360 degree feedback, Valle & Bozeman, 2002). One reason is that teachers usually cannot observe all activities of their principals. Teachers' perceptions may therefore be strongly influenced by those observations they made when being in direct contact with their principals or their deputies, for example, during general assemblies, face-to-face conversations, or collaborative work in panels. Similarly, teachers also vary in their perceptions, which is evident in the interrater reliabilities of teachers perceptions of previously used evidence, which did not exceed ICC1 = .50 (used external inspections) and was sometimes as low as ICC1 = .22 (used internal feedback). It would be interesting to know if expected positive outcomes of EBSMgt are more strongly affected by principals' perceptions' of their evidence-related activities, by teachers' perceptions of these activities, or by the shared perceptions of teachers and principals. Consider, for example, the case of using evidence to support teachers in their work. The literature is quite clear that *perceived availability* of support benefits employees' motivation and health, whereas *received* support has been repeatedly shown to relate to distress (cf. Gleason & Iida, 2015). In a similar vein, evidence-oriented decision-making of principals that is not perceived as such might undermine teachers' motivation and health, whereas teachers' perceptions of principals' evidence-oriented decision-making is beneficial to them. Furthermore, organizational climate research has always been concerned with the question if individual's perceptions or the shared perceptions of all employees are more important for outcomes at the individual level (e.g., Schneider, Bowen, Erhart, & Holcombe, 2000). In the present study, we decided to focus on aggregate teachers' perceptions and averages of teachers' and principals' perceptions of EBSMgt for analysing different types of EBSMgt. The results might have been different when using different approaches. More detailed comparisons of different approaches, however, were beyond the scope of the present study, and future research has to address these issues to better understand how the effects of EBSMgt unfold.

We should note that ICCs for principals suffered from the fact that only a few principals and their deputies were sampled per school (1.89 persons per school on average). This is the normal situation in Germany, where in most school types school management teams rarely comprise of more than four persons. Studies in other countries with schools having larger management teams might be better suited to analyze whether interrater reliability exists among principals and their deputies. Furthermore, the schools in our sample included all types of schools

in Germany, but they were all sampled within a single federal state (Rhineland-Palatinate). In Germany, schools are governed by federal states, which makes schools within a federal state more similar to each other than schools belonging to different federal states. Moreover, they were all public schools. Private schools, in particular those that work for profit, may depend much more on being successful, which could be a driver for EBSMgt. Although for-profit schools are not yet too common in Germany, they are more prevalent in other countries. Including private schools and schools representing a broader regional scope in future studies could increase the variance in evidence orientations among schools, which should also result in higher ICCs, indicating that EBSMgt can be reliably assessed at the school level.

Among individual teachers, the three scales were substantially correlated ($-.33 \leq r \leq .61$). Among individual principals, these correlations were much smaller ($-.17 \leq r \leq .37$). Aggregated school level data yielded almost identical results. The strongest correlation emerged for teachers' assessments of IE and EE, which could threaten discriminant validity. Perhaps, teachers' ratings were contaminated by a halo effect. It could well be that slight indications of principals' high evidence-oriented or low evidence substitute-oriented behavior was regarded as an indicator that the respective opposite behavior is not shown by their principals.

Despite the substantial correlations of the scales among teachers, using teacher data and further variables established convergent and discriminant validity of the three scales at the individual level of analysis. Proposed antecedents (i.e., previous use of information that resulted from evidence-oriented activities) and most proposed consequences (i.e., the rated usefulness of this information) were related to IE, EE, and ES as expected. Interestingly, on the one hand, participating in external inspections and using information resulting from such inspections spread its effects to all three dimensions, that is, it increased IE and EE and it decreased ES. Conversely, high IE and EE and low ES simultaneously impacted on the rated usefulness of external inspections. Thus, participating in external inspections could be a door opener to increase EBSMgt and to further regard EBSMgt as really useful.

In line with van Ackeren et al. (2013), we suggest using the term EBSMgt when the combination of high evidence *orientation* and low ES is present. Based on the work by Stumm et al. (2010), van Ackeren et al. (2013) distinguished four types of information seeking behavior in schools, focusing more or less strongly on evidence and on evidence substitutes. In our study, most schools are characterized by *EBSMgt* (31 out of 157 schools). They have high scores on IE and EE and low scores on ES. The second largest group, however, is *evidence-averted school management* (29 schools). These schools have low scores on IE and EE and high scores on ES. The third largest group comprises of schools characterized by *sclerotic school management* (25 schools). These schools have low scores on all three scales. These schools seem to be either not interested in any evidence that could be used for school development, or they do not engage much in school development. All other combinations of the three scales do not represent larger portions of the entire sample of schools. In particular, we could not identify a substantial

group of schools with the *volatile school management* type, which was reported by van Ackeren et al. (2013). Probably, this was due to the fact that we simultaneously used teachers' and principals' perceptions and gave them equal weight.

Although the result that the largest group of schools is characterized by EBSMgt, two critical management styles (evidence-averted and sclerotic school management) together represent more than one third of the sample. Even though the cut-off points to dichotomize the sample on each of the three scales were empirically determined and can, therefore, be challenged, with the exceptions of IE (3.81) the cut-off values were close to the scale midpoint (2.79 for external evidence and 2.89 for substitute orientation), which reflect medium agreement that the listed activities are present. Thus, we believe, that indeed for one third of the schools an increase in EBSMgt is desperately needed.

4.1 Limitations

Despite its strength, our study and its results could suffer from some limitations. As already mentioned, our sample of schools, principals and teachers is likely to be not fully representative due to nonresponse and restriction to major school types. In addition, we only sampled public schools within a single state of Germany. In particular, we cannot claim that the percentages of the different EBSMgt types obtained in our study are representative. Increasing the variability by including private schools and schools from a more diverse regional background would also be helpful to further validate the concept of EBSMgt and its measurement tool.

Our attempt was to establish convergent and discriminant validity by using proposed antecedents and consequences. For this purpose, we believe that using only cross-sectional data was an acceptable approach. However, obviously, this renders any causal explanation problematic, and the research model that guided our research (see Figure 1) cannot be regarded as a process model of EBSMgt that was confirmed in the present study. Although we believe that participating in evidence-based activities and using information resulting from such activities indeed leads to a climate for evidence which, in turn, makes people believe that using evidence is really useful for school development, longitudinal studies are required to confirm such a model with higher fidelity than our study can do. Thus, above and beyond establishing convergent and discriminant validity, longitudinal research is required to better establish what drives EBSMgt and what its outcomes are.

Some psychometric properties of the developed scales were not really satisfactory. As already discussed, it might or it might not be possible to increase reliability of the ES scale. A larger set of items addressing substitute orientation may also unfold to more than a single factor. Future research is certainly needed in this respect, too.

One could argue that the measurement tool for EBSMgt developed in the present study lack some content validity as it does not include the specific evidence-oriented activities that have become prominent, such as external school evaluations

like PISA or IGLU/PIRLS. However, we do not believe this is a real shortcoming. Rather, it was an intended feature of the scales. Particular measures are only available to particular schools, such as PISA for high schools or IGLU for primary schools. Thus, comparing the extent of EBSMgt across schools of different types or located in different areas would be difficult if scales using different sets of items were applied. The three scales developed in this paper, that is, EE, IE, and ES, which together characterize the type of EBSMgt, do allow for such comparisons.

4.2 Conclusions

There has been little evidence on EBMgt in general (Reay et al., 2009) and on EBSMgt in particular. We believe that the present study represents an important starting point to empirically investigate the concept of EBSMgt as it provided three scales, which allow assessing EBSMgt on a relative abstract level. This provides a foundation for future studies, which will hopefully yield further refinements and more elaborated theoretical models describing the causes and consequences of EBSMgt. Empirically establishing the consequences of EBSMgt is of pivotal importance for the next couple of years. Indeed, we strongly believe (but do now yet know) that EBSMgt is a successful strategy for improving the schooling of those who follow, and that the success of EBSMgt extends well above grades and knowledge leading to “fully functioning persons”.

Acknowledgements

This present paper was supported by a collaborative research grant provided by the German Federal Ministry of Education and Research (BMBF) to the University of Mainz and the University of Duisburg-Essen (01 JG 1010 A & B).

References

Ackeren, I. van, Binnewies, C., Clausen, M., Demski, D., Dormann, C., Koch, A., Laier, B., Preisendörfer, P., Preuße, D., Rosenbusch, C., Schmidt, U., Stump, M., & Zlatkin-Troitschanskaia, O. (2013). Welche Wissensbestände nutzen Schulen im Kontext von Schulentwicklung? Theoretische Konzepte und erste Befunde des EviS-Verbundprojektes im Überblick [What kind of evidence do schools use for school development: Overview of theoretical concepts and first results of the EviS-collaborative research project]. In I. van Ackeren, M. Heinricht, & F. Thiel (Eds.), *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund* (Die Deutsche Schule. Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis, Beiheft 12, pp. 51–73). Münster, Germany: Waxmann.

Ackeren, I. van, Zlatkin-Troitschanskaia, O., Binnewies, C., Clausen, M., Dormann, C., Preisendörfer, P., Rosenbusch, C., & Schmidt, U. (2011). Evidenzbasierte

Schulentwicklung: Ein Forschungsüberblick aus interdisziplinärer Perspektive [Evidence-based school development: An interdisciplinary review of research]. *Die Deutsche Schule, 103*(2), 170–184.

Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Battistelli, A., Montani, F., & Odoardi, C. (2013). The impact of feedback from job and task autonomy in the relationship between dispositional resistance to change and innovative work behaviour. *European Journal of Work and Organizational Psychology, 22*(1), 26–41. doi:10.1080/1359432X.2011.616653

BR – Bayerischer Rundfunk. (2013). *Ermittlungen wegen geschönter Abitur-Noten* [Investigation due to flattering high school graduations grades]. Retrieved from http://www.br.de/nachrichten/oberfranken/coburg-abi-schummelvorwuerfe-100.html

Briner, R. B. (2007). *Is HRM evidence-based and does it matter?* (Opinion Paper OP6). Retrieved from http://www.employment-studies.co.uk/pdflibrary/op6.pdf

Briner, R. B., Denyer, D., & Rousseau, D. M. (2009). Evidence-based management: Concept clean-up time? *Academy of Management Perspectives, 23*(4), 19–32. doi: 10.5465/amp.200945590138

Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I-O psychology: Not there yet. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4*(1), 3–22. doi: 10.1111/j.1754-9434.2010.01287.x

Externbrink, K., & Dormann, C. (2015). Führen und Entscheiden: Evidence-based Management [Leading and deciding: Evidence-based management]. In J. Felfe (Ed.), *Trends der psychologischen Führungsforschung* [Current trends in psychological leadership research] (pp. 429–439). Goettingen, Germany: Hogrefe.

Gleason, M. E. J., & Iida, M. (2015). Social support. In M. Mikulincer, P. R. Shaver, J. A. Simpson, & J. F. Dovidio (Eds.), *APA handbooks in psychology. APA handbook of personality and social psychology* (Vol. 3, pp. 351–370). Washington, DC: American Psychological Association. doi: 10.1037/14344-013

Gruber, H., Harteis, C., & Rehrl, M. (2006). Professional Learning: Erfahrung als Grundlage von Handlungskompetenz [Experience as a foundation of action compentence]. *Bildung und Erziehung, 59*(2), 193–203.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*(2), 219–229.

Lehmann, C. (2013, July 16). *Coburger Abitur-Ergebnisse vom Direktor geschönt?* [Graduation results in Coburg cooked by the school principal?]. *inFranken.de*. Retrieved from http://www.infranken.de/regional/coburg/Casimirianum-Coburg-abitur-Coburger-Abitur-Ergebnisse-vom-Direktor-geschoent;art214,476066

Maier, U., & Kuper, H. (2012). Vergleichsarbeiten als Instrument der Qualitätsentwicklung an Schulen. Überblick zum Forschungsstand [Parallel exams as a tool for quality-oriented development in schools. Review of research]. *Die Deutsche Schule, 104*(1), 88–99.

Pfeffer, J., & Sutton, R. (2000). *The knowing-doing gap. How smart companies turn knowledge into action*. Boston, MA: Harvard Business School Press.

Pfeffer, J., & Sutton, R. (2006). *Hard facts, dangerous half-truths and total nonsense*. Boston, MA: Harvard Business School Press.

Posch, P. (2009). Zur schulpraktischen Nutzung von Daten: Konzepte, Strategien, Erfahrungen [Abouth the practical use of data in schools: Concepts, strategies, experiences]. *Die Deutsche Schule, 101*(2), 119–135.

Reay, T., Berta, W., & Kohn, K. (2009). What's the evidence on evidence-based management? *Academy of Management Perspectives, 23*(4), 5–18. doi: 10.5465/AMP.2009.45590137

Rousseau, D. (2006). Is there such thing as "evidence-based management"? *Academy of Management Review, 31*(2), 256–269. doi: 10.5465/AMR.2006.20208679

Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management, 36*(4), 941–973. doi: 10.1177/0149206309350084

Schneider, B., Bowen, D. E., Ehrhart, M. G., & Holcombe, K. M. (2000). The climate for service: Evolution of a construct. In N. M. Ashkanasy, C. Wilderom, & M. F. Peterson (Eds.), *Handbook of organizational culture and climate* (pp. 21–36). Newbury Park, CA: Sage.

Stumm, S., Mohr, N., & Dormann, C. (2010, July). *Evidence based management – An effective approach for public agencies?* Presentation at the 27th International Congress of Applied Psychology, 11.–16.07.2010, Melbourne, Australia.

Valle, M., & Bozeman, D. P. (2002). Interrater agreement on employees' job performance: Review and directions. *Psychological Reports, 90*(3), 975–985. doi: 10.2466/PR0.90.3.975-985

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk, 10*(3), 295–308.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods, 12*(4), 695–719. doi: 10.1177/1094428108327450