

Bea Harazd, Stefanie van Ophuysen, Julia Gerick & Mario Gieske

How accurately do principals evaluate the strain experience of their teaching staff?

Abstract

Within the framework of expanded school autonomy design in Germany, school administrations have to deal with more extensive management responsibilities which include, among other things, a more health-oriented school design. But how do principals and teaching staff assess the potential strain factors in the school environment? How well do their assessment-scores correlate? To answer these questions, about 1,600 teachers from 45 schools in North Rhine-Westphalia were asked in an online questionnaire to evaluate the strain they experienced in six different areas of the school context. Their principals (n = 45) also evaluated the strain experience of their respective staffs. Overall, everyday occupational life at school is perceived as being slightly straining. However, there are clear differences between various areas of activity. While about 60 % of the data from principals stand in line with the data submitted by teachers, principals tend to underestimate the strain experienced by their staff. In particular, this effect is due to misjudgments in the area of conference and teamwork. The consequences of this lack of judgment accuracy for the fulfillment of a 'health-oriented school design' are discussed as well as the question of how an improved evaluation can be reached in the interest of principals' professional management.

Dr. Bea Harazd, University of Münster, Georgskommende 33, 48143 Münster, Germany
e-mail: bea.harazd@uni-muenster.de

Prof. Dr. Stefanie van Ophuysen, University of Münster, Georgskommende 33, 48143 Münster, Germany
e-mail: vanophuysen@uni-muenster.de

Dipl.-Päd. Julia Gerick (corresponding author), Institute for School Development Research, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany
e-mail: gerick@ifs.tu-dortmund.de

Dipl. Reha.-Päd. Mario Gieske, TU Dortmund University, Hohe Str. 141, 44139 Dortmund, Germany
e-mail: mario.gieske@uni-dortmund.de

Keywords

Diagnostic competence; Principal; Stress; Health Promotion

Wie akkurat schätzt die Schulleitung das Beanspruchungserleben des Kollegiums ein?

Zusammenfassung

In Deutschland kommen im Rahmen der erweiterten Gestaltungsautonomie von Schulen auf die Leitung einer Schule umfangreichere Leitungsaufgaben zu, u. a. auch eine gesundheitsförderliche Schulgestaltung. Aber wie beurteilen Schulleiter und Lehrkräfte die möglichen Belastungsfaktoren im schulischen Umfeld? Wie gut stimmen die Einschätzungen der Befragtengruppen überein? Zur Beantwortung dieser Fragen wurden in NRW rund 1600 Lehrkräfte aus 45 Schulen gebeten, die von ihnen erlebte Beanspruchung in sechs verschiedenen Bereichen des schulischen Kontextes in einem Online-Fragebogen einzuschätzen. Auch ihre Schulleiter (n = 45) schätzten das Beanspruchungserleben des jeweiligen Kollegiums ein. Insgesamt wird der berufliche Alltag in der Schule als leicht belastend wahrgenommen, wobei es deutliche Unterschiede zwischen verschiedenen Tätigkeitsbereichen gibt. Wenngleich rund 60 % der Schulleiterangaben mit den Lehrerangaben übereinstimmen, deutet sich insgesamt dennoch eine leichte Unterschätzung der Lehrerbeanspruchung durch den Schulleiter an. Dies ist insbesondere auf Fehleinschätzungen im Bereich der Konferenz- und Teamarbeit zurückzuführen. Die Konsequenzen der eher geringen Urteilsgenauigkeit für die Erfüllung der Schulleitertätigkeit „gesundheitsförderliche Schulgestaltung“ werden ebenso diskutiert wie die Frage, wie eine Verbesserung der Einschätzung im Dienste eines professionellen Schulleiterhandelns erreicht werden kann.

Schlagworte

Diagnostische Kompetenz; Schulleitung; Belastung; Gesundheitsförderung

1. Introduction

The health of and strain on teachers is a current area of concern in Germany and an ongoing topic of debate, both in the public sphere and in empirical research. One reason for this is the high number of early retirements on grounds of disability. From 1993 to 2001, between 80 % and 93 % of German retired teachers retired before the legal standard retirement age. Between 34 % and 64 % of these retirees stopped working due to illness (Federal Statistical Office of Germany, 2008, p.100). This quota is above that of comparable professions in the public service (Federal Statistical Office of Germany, 2008, p.97 et seq.). One central reason for this high rate of early retirees is the teachers' experience of stress. According to several em-

irical studies, a substantial proportion of teachers – not only in Germany – perceive their job as stressful (Borg, Riding, & Falzon, 1991; DeFrank & Stroup, 1989; Griffith, Steptoe, & Copley, 1999; Rudow, 1994; van Dick, 2006), and it is concluded that teaching is one of the most stressful professions (Kyriacou, 2001).

Even if stress does not lead to illness in all cases, it is negatively related with health – in the sense of its broader definition, which was promoted by the World Health Organization’s 1986 Ottawa Charter. In this definition, health is not only understood as the absence of illness, but rather as a “state of complete physical, mental and social well-being” (WHO, 1986, p. 1).

The growing significance of health and health promotion in German schools is elucidated by the fact that in several German federal states (e.g., Brandenburg, Lower Saxony, Hamburg, Baden-Wuerttemberg, Berlin, Hesse), policy guidelines for school quality have been extended with the aspect of health-oriented working conditions or by the aspect of working environment design. Furthermore, the health of students and teaching staff has become a part of the school inspection in North Rhine-Westphalia. The existence of a concept “for an active preventative health care for teachers and other employees” (Ministry of School and Further Education of the Federal State of North Rhine-Westphalia, 2006, p.11) is seen as a characteristic of quality. This is in line with empirical research supporting the principals’ importance for teacher health. If the principal is perceived as supportive, teachers indicate a higher mental well-being (e.g., Matthews, Cottington, Talbot, Kuller, & Siegel, 1987) as well as less stress (Blase, Dedrick, & Strathe, 1986) and burnout (Burke, Greenglass, & Schwarzer, 1996; Leithwood, Menzies, Jantzi, & Leithwood, 1999).

To ensure this aspect of quality at school, the Education Act of the Federal State of North Rhine-Westphalia (2005) established occupational safety and health protection as management responsibilities, alongside personnel management and development and the development of organization and instruction (§59 par. 8). But how can principals fulfill this demanding task?

A first focus might be on interventions concerning objective working conditions and situations on the one hand and the individual teacher’s strengths and weaknesses on the other hand. Principals should be able to assess these characteristics as accurately as possible in order to respond with appropriate and specific interventions (e.g., collective strategies to deal with disciplinary problems and classroom disturbances, changes in workplace design, training on team work).

However, transactional stress models (esp. Lazarus, 1966) postulate that strain occurs if person-specific options for action are perceived as insufficient for coping with situation-specific demands. Therefore, next to objective personnel and situational conditions, it is important to take the individual interpretations, constructions and assessments related to a given situation into account. In order to establish an effective health concept, principals should be able to detect these individual perceptions of a given situation, too.

Principals face a major challenge with regard to this task, the fulfillment of which requires a high level of empathy and an ability for social perception. How well they perform on this task has not been examined empirically so far.

1.1 Facets of strain in schools for teaching staff

Which facets of their occupation do teachers perceive as potentially straining? Which are perceived as being particularly negative? These questions have been addressed in a whole series of empirical studies. In a study by Häbler and Kunz (1985), students' lack of motivation and concentration (51.8 %) as well as preparation and evaluation of lessons (44.8 %) were experienced as being particularly straining. A survey of about 500 teachers from North Rhine-Westphalia and Hessen demonstrated that oversized classes as well as a lack of motivation and concentration represent the most severe strain factors (van Dick, 2006).

However, direct comparison of empirical results is difficult, because both the questions and the response formats differ significantly between the studies. In response formats both frequency and intensity were asked for. Likewise, content is not comparable, because in different studies different conditions were examined. In most studies on strain experience, lists of specific situational conditions are provided and teachers are asked to evaluate the potential strain they experience with regard to each of these conditions. Considering school-specific as well as context-specific working conditions is a major advantage of this individualized assessing of strain. However, in addition to the question of comparability there is also the question of completeness: Are all relevant factors included in each study? To answer this question, a system of generic categories which systematically cover the field of possible strain factors is necessary. For example, Kramis-Aebischer (1995) suggests a formal model of strain analysis that incorporates strain at system, school and individual levels.

Rudow (1994) suggests separating *neutral* working conditions such as (school) organizational conditions, working environment conditions and social and cultural conditions from *negative* working conditions such as discipline problems or bullying. However, this *a priori* systemization of potential stressors is only of limited use as a categorical framework because it presupposes that negative working conditions can be identified in advance. This is a somewhat implausible approach, particularly within the context of a transactional model that focuses on the interaction of situational demands and personal resources. Such a categorical separation may be useful after the analysis of empirical materials, working out precisely which situational conditions, irrespective of personnel resources, are perceived as strain (for further details see Gieske, Harazd, & Gerick, 2010; Harazd & Gieske, 2009).

More powerful content-related proposals for the categorization of stressors and of straining working conditions originate from "industrial psychopathology". Udris and Frese (1988) distinguish between stressors caused by (a) *work related tasks* (e.g., aspiration level), (b) *physical* stressors (e.g., heat, noise, dirt), (c) stressors of

a *temporal* nature (e.g., shift and night work) and (d) stressors from *social* and *organizational* situations (e.g., role ambiguity, working atmosphere).

In a very similar manner, McGrath (1981) distinguishes between five work related stress factors: (a) the task itself (e.g., monotony), (b) work organization (e.g., time pressure), (c) physical conditions (e.g., noise), (d) social conditions (e.g., unfairness) and (e) the general organizational conditions (e.g., status and acknowledgment).

This differentiation by McGrath (1981) seems to be rather well suited to categorizing potential stressors in the teaching profession at school level. For example, workplace characteristics proposed by Rudow (1994) or Schaarschmidt (2005) can be arranged clearly according to this category scheme. Nevertheless, we propose two changes: Firstly, characteristics such as noise, dirt or heat fade into the background when compared to the demands of the job itself, so they need not be regarded as an independent aspect. Secondly, work-related tasks in the teaching profession are highly complex and varied, and need additional differentiation. Work-related tasks referred to by Rudow (1994) and Schaarschmidt (2005) comprise – besides teaching as the intrinsic core concern – duties among teaching staff (conferences, team and committee work) as well as innovation and school development (implementation of reforms).

In order to assess the working conditions of teaching staff, we propose a six category system based on that of McGrath, whereby alongside the three facets of work-related tasks (teaching, teamwork, and reform and innovation), the areas of social relationships, organizational conditions and organization of work at school are incorporated.

1.2 Accuracy of judgment in the school context

The question of principals' accuracy in perceiving the strain on teaching staff is not only related to existing research on teachers' strain; empirical findings from research on the topic of "diagnostic competence" are also relevant. There are a number of insightful findings about the accuracy of teacher evaluation with regard to the individual learning efficiency or aptitude of their students (for a summary see Hoge & Coladarci, 1989; Spinath, 2005; van Ophuysen, 2010). However, these studies differ from the question raised here in three respects. Firstly, previous studies involve the evaluation of students by their teachers. Secondly, the studies predominantly examine the areas of learning and achievement or further related individual characteristics as motivation or intelligence. Thirdly, diagnoses are generally considered at individual level (however, see van Ophuysen, 2009). In the current study, however, the focus is on (a) the evaluation of teachers by their principals with regards to (b) their strain experienced due to their profession on (c) the level of the staff as a group. Both the evaluation content considered here and the focus on group perspective raise methodological problems, which are discussed below.

The accuracy of an evaluation is usually determined by the comparison of “true” and “diagnosed” characteristic values. For this, Schrader and Helmke (1987) distinguish between a level component (correct evaluation of the mean characteristic value in the target group), a differentiation component (correct evaluation of the distribution of characteristic values) and a rank component (correct evaluation of the ranking order of characteristic values).

Such analyses require the determination of a “true” value of the characteristic under investigation. For example, when dealing with the teachers’ accuracy in assessing students’ competences, reliable measures of test performance or achievement-related self-evaluations are chosen as “true” scores (Hosenfeld, Helmke, & Schrader, 2002; Spinath, 2005; Wild, 1991). In analogy to this approach, the experience of strain can be evaluated individually by teachers with the aid of a strain index including a variety of school-specific working conditions. An aggregated mean value of the teaching staff can be utilized as a reference value. However, in comparison to student performance, there is a lack of established testing procedures with assured quality criteria for measuring teachers’ strain at school. Accordingly, reference values might have potentially lower reliability.

Next to the operationalization of the “true” reference value for teachers’ strain, the method for assessing the principals’ “diagnosis” is of high importance in estimating diagnostic accuracy, too. On the one hand, principals and teaching staff can answer parallelized items. Thus, the understanding of the construct to be measured is similar for both groups of respondents and the scale level of both measurements coincides. In this case, the data from principals and (aggregated) data from teachers can be directly compared. For practical reasons, however, the number of items often has to be kept low when surveying experts. Therefore, instead of asking for a list of specific situations, only single items are utilized representing rather global ratings, e.g., above average/below average or problematic/unproblematic (German PISA Consortium, 2001; Spinath, 2005; van Ophuysen, 2009). In this case, the ordinal measurement value of a single item must be compared to the reference score, which is assessed on a metric scale.

Overall, it should be emphasized that reviewing accuracy is essentially interconnected with determining a correct reference value and using parallel instruments of data collection. Particularly in the case of characteristics for which no standardized testing procedures with standard values exist, rather unreliable scores have to be expected, and this vagueness must be taken into account in the interpretation.

1.3 Research questions

The central aim of our research is to determine how successful principals are at evaluating the (average) level of strain experienced by their teachers. For two reasons we focus on the aggregated evaluation of the entire teaching staff. On the one hand, it is plausible that principals decide about the initiation of health measures on the basis of considerations concerning the welfare and requirements of

the teaching staff as a whole instead of focusing on the single teachers' demands. In addition to this reason with regard to context, practical and legal considerations (data protection laws) had to be taken into account.

The central issue of judgment accuracy is preceded by the descriptive analysis of levels of strain in various areas as reported by the teachers. Which areas are experienced by teaching staff in our sample as being particularly straining? Which ones potentially relieve stress? On the other hand, we will have a look at the principals' estimation of strain experienced by their staff with regards to different areas.

2. Methods

2.1 Data collection procedures

This study was conducted within the framework of a three year research project "Leadership Concepts in Good and Healthy Schools", which is concerned with the relationship between leadership behavior, school quality and teacher health. Data collection from principals and staff was carried out by means of a standardized online questionnaire in March and April 2008. Principals were requested in advance, both in writing and by telephone, to participate. As a rule, principals agreed to take part after consulting with their entire staff. Completion of the extensive online questionnaire took approximately 30 to 45 minutes. Participant anonymity was guaranteed through the allocation of individual access codes. A telephone hotline was available for assistance with technical and content-related queries. At the request of principals, a paper-pencil version of the questionnaire was administered at some schools instead.

2.2 Sample

From 211 schools in North Rhine-Westphalia that were invited to take part in the study, principals and teachers of 125 schools (33 elementary schools, 24 high schools, 23 comprehensive schools and 45 vocational colleges) agreed to participate. However, response rates from the teaching staffs of the participant schools varied considerably (min = 2 %, max = 100 %) with an average response rate of 40 %. In order to compare the principals' evaluation with the (average) perception of strain in the teaching staff, we considered it inappropriate to include schools in which only a small number of teachers had completed the questionnaire. Accordingly, only the 45 schools in which at least 50 % (average 72 %) of the teaching staff participated were included in the analysis. As it becomes clear from this description, our sample is not a probability sample and therefore, we cannot assume the sample to be representative of schools in NRW.

Table 1: Investigation sample

Type of school	> 50 %		Response teaching staff		
	Principals	Teachers	Average	Min	Max
Elementary School	21	379	18	14	36
Gymnasium (High School)	5	221	44	63	86
Comprehensive School	3	216	72	71	96
Vocational College	16	827	52	58	91
Total	45	1,643			

In addition to the 45 principals, 1,643 teachers were surveyed within these schools. The frequency distribution for school types is reported in Table 1. In comparison with the entire teacher population in North Rhine-Westphalia (Ministry of School and Further Education of the Federal State of North Rhine-Westphalia, 2007), teachers in our subsample are slightly younger (45.2 years old vs. 47.8 years old) and the percentage of female teachers is somewhat lower (58.3 % vs. 60.7 %).

2.3 Instruments

Within the framework of the extensive online questionnaire, particular items were introduced for the evaluation of strain experienced by teachers. Although principals and teaching staff were surveyed on the same subject areas, the instruments used are not completely parallelized.

Questionnaires for teachers. Teachers read 16 neutrally formulated items describing aspects of their average working day (e.g., social climate, further training possibilities, student behavior; see Table 2). For each item they were asked to indicate on a 7-point bipolar rating scale (from -3 = very straining to +3 = very relieving) how straining or relieving they perceived these to be. The items were sub-divided into the aforementioned six areas of strain based on the model of Udriș and Frese (1988). Each category was represented by one to five items. For each category, a mean item score was computed as an area-specific index value of the strain experienced. Relieving areas are represented by a positive value and straining areas by a negative value. Furthermore, an overall index of strain was computed as the average over the area-specific means.

Questionnaires for principals. The principals' task was to evaluate globally the severity of strain or relief experienced by their teaching staff for each of the six areas of strain. As with the teachers' questionnaire, the structured response format was 7-point (from -3 = very straining, to 0 = neutral (neither/nor) through to +3 = very relieving).

Table 2: Instruments of data collection for principals and teachers

Principal items	Teacher items
General organizational conditions of the school (e.g., acknowledgment and remuneration, opportunities for codetermination, promotion of further training)	Handling of acknowledgment and remuneration Promotion of further training Opportunities for codetermination
Organization of work at school (e.g., organization of lessons, working hours, organizational structure, spatial and material equipment)	Organization of lessons Organizational structure of the school Spatial and material equipment in the school
Social conditions (e.g., working climate within the teaching staff, collaboration with parents and students)	Collaboration between parents and teaching staff Working climate among the teaching staff
Conference and team levels (e.g., committee and conference work)	Committee and conference work
Teaching conditions (e.g., class work, preparation and follow-up, corrections, student behavior in class)	Volume of the lessons Correction of class work Preparation and follow-up of classes Number of students Behavior of students in class
Current Reforms (e.g., implementation of grades for study habits and social behavior, centralized examinations)	Implementation of grades for study habits and social behavior Centralized examinations

Nevertheless, in order to ensure that the indices calculated from the teacher data are broadly compatible with the principals’ evaluations regarding content, the areas of strain were exemplified in the principals’ questionnaire via the individual conditions evaluated by the teachers. An overall strain index was also formed here by calculation of the mean value across the six areas.

2.4 Data analysis

In a first step, data was analyzed in a descriptive way, separately for both groups (teachers and principals). Aggregated teacher data serve as an indicator for the actual status quo at each school. The principals’ data represent the diagnosis of these conditions. Analyses are based on six area-specific indices of strain as well as on the overall mean. Measures of central tendency and of dispersion are reported and paired sample *t*-tests are computed for comparing teachers’ and principals’ mean scores. This analysis allows for the assessment of the *average accuracy of judgment* of all principals throughout and clarifies the question: “How well do the principals’ evaluations correspond with the teaching staffs’ evaluations with regard to central tendency and dispersion?”

Furthermore, by means of non-parametric correlation analysis, the rank order of school-specific values of teachers and principals is compared. Do principals tend to report high strain on their teachers if teacher data reveal high strain experience?

An informative way of coupling the data is to compute school-specific deviances (aggregated teacher score minus principal score) and their absolute values. The mean deviance score provides information about the level component of accuracy. A positive value represents an overestimation of strain by principals while a negative value represents an underestimation. The mean absolute deviance provides information about the average extent of deviation – regardless of its direction. High values indicate strong differences in the evaluation of the teachers' strain between principals and the staff itself.

Although mean deviance is a sensible measure for group mean differences, this measure does not take teachers' individual perceptions into account. Perceptions of the working conditions vary within each school. A high internal differentiation within a teaching staff (i.e., a high variation in the values of individual teachers) makes it particularly difficult for principals to render an overall judgment. In order to take this variation of individual teacher data within schools into account, a tolerance interval of plus/minus one standard deviation is placed on the mean value in each school. Assuming normal distribution, 68.3 % of data lie within this interval. If a principal's value lies below the lower interval boundary or above the upper interval boundary, his evaluation is categorized as an under- or overestimation, respectively. If the principal's evaluation falls within the tolerance interval, it is interpreted as a match.

Finally, we tested if accuracy of principal's evaluation depends on the percentage of teacher feedback. The sample of 45 schools was divided into two groups by median split (22 schools with a response rate of less than 65 %, 23 schools with a response rate above 65 %), and mean total deviances between these two sub-samples were compared by means of an independent *t*-test.

3. Results

3.1 Area-specific strain factors from the perspectives of teachers and principals

Looking at the data from teachers, it is revealed that work organization in general and concerning instruction is perceived as neither straining nor relieving. Work itself – teaching, conferences and most of all implementing current reforms – is seen as a source of strain. On the other hand, social conditions at work are reported to be relieving. Standard deviations indicate that variation in perception is comparable for all six aspects under investigation. Despite noteworthy variation in the school wise average scores for strain by implementation of current reforms, all means are negative, indicating that this aspect of the teachers' work is experienced stressful in all schools.

Along general lines, principals' evaluations mirror their teaching staffs' view: In three of six aspects no significant difference in evaluation can be revealed, and al-

though means differ statistically significant, reform implementation is perceived as straining and social conditions are perceived as relieving by teachers as well as by their principals. However, teachers report conference and team work to be a stressful part of their job, while principals assume that this aspect is perceived as relieving by their teaching staff. Because of the small sample size ($n = 45$), all statistically significant differences ($p < .05$) exceed an effect size of $d = 0.40$ (Cohen, 1988, Chapter 2.4.5).

The variance in principals' data is somewhat wider than the distribution of teachers' evaluations. However, this is attributable to the fact that the teacher data were calculated as mean values of the individual scores.

Mean values and standard deviations as well as results of the t -tests are summarized in Table 3.

Table 3: Principal and teacher evaluations (mean values, standard deviation, difference and results of the t -test for group differences)

Area of Strain	Principals			Teaching staff			t-Test		Cohen's d
	Mean	SD	Min/Max	Mean	SD	Min/Max	t	p	
1. General organizational conditions of the school (e.g., acknowledgment and remuneration, opportunities for code-termination, promotion of further training)	0.56	1.47	-3/3	0.32	0.51	-1.35/1.51	-1.16	.25	0.22
2. Organization of work at school (e.g., organization of lessons, working hours, organizational structure, spatial and material equipment)	0.27	1.71	-3/3	0.02	0.58	-0.96/1.71	-1.08	.28	0.19
3. Social conditions (e.g., working climate among the teaching staff, collaboration with parents and students)	1.24	1.48	-3/3	0.76	0.42	-0.25/1.62	-2.30	.03	0.45
4. Conference and team levels (e.g., committee and conference work)	0.47	1.44	-3/3	-0.77	0.43	-1.73/0.50	-5.96	< .01	1.17
5. Teaching conditions (e.g., teaching, preparation and follow up, corrections, student behavior in class)	-0.89	1.43	-3/2	-0.79	0.32	-1.45/0.12	0.49	.62	0.10
6. Current Reforms (e.g., implementation of grades for study habits and social behavior, centralized examinations)	-1.93	1.23	-3/3	-1.35	0.29	-0.75/-2.21	3.13	< .01	0.66

Integrating all area-specific evaluations into one global sum score provides a reliable measure of perceived stress for both principals and teachers ($\alpha_{\text{principal}} = .76$, $\alpha_{\text{teacher}} = .75$). For this global score, a mean value of -0.30 ($SD = 0.30$) results among the teachers, while the principals' score is -0.05 ($SD = 0.90$). The t -test for depend-

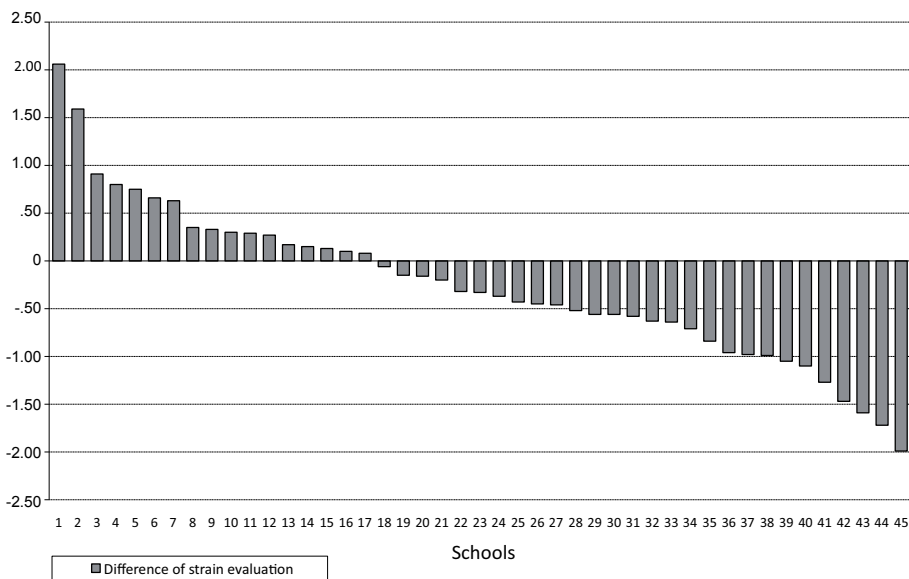
ent samples indicates the difference between these mean values as statistically significant ($t = -2.08$; $df = 44$; $p = .04$).

3.3 Global strain scores in pair-wise comparison

The comparison of distribution parameters presented thus far (mean values and standard deviations) from principals' data and aggregated teacher data for each school enables the assessment of the *mean* accuracy of judgment across all principals.

In the next step of the analysis, principal and teacher data are coupled school by school to examine rank order consistency by means of a non-parametric correlation analysis.

Figure 1: Evaluation of strain among the teaching staff; comparison of principals' data with the aggregated teachers' data per school ($n = 45$)¹



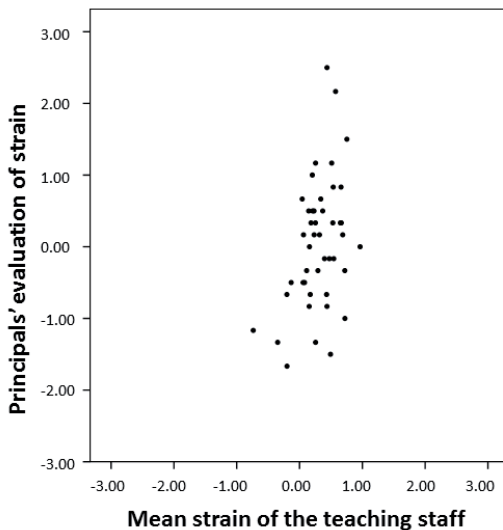
The linear relation between rank data from principals and teachers is relatively small at $r = .33$ ($p = .027$). However, in interpreting the correlation, it should be noted that the aggregated teacher data lie relatively close to the mean value, revealing only a slight dispersion; this becomes clear both in Figure 1 and in the standard deviations in Table 3.

1 7-point response format for both sample groups (3 = very relieving; 2 = relieving; 1 = somewhat relieving; 0 = neither/nor; -1 = somewhat straining; -2 = straining; -3 = very straining)

The inspection of the difference values and their absolute values allows for further assertions about the accuracy of judgment. The graphic representation of the difference values, ordered by their sizes, underlines the results hitherto with two observations: First, it is once again evident that overall an underestimation of strain occurs more frequently than an overestimation indicating a weak accuracy with regard to the level component. At the same time, it is evident that both forms of misjudgment exist, from which poor accuracy regarding the rank components can be inferred (see Figure 2).

The mean of the absolute deviations is 0.68. Thus, on average the principals fall short of the evaluations of their teaching staff by 0.68 points (regardless of direction). To check the effect of school wise response rate on the accuracy of the principals' assumptions about their teaching staffs' evaluations, the sample was divided into two groups by using median split (Median = 65 %). No significant difference in accuracy (operationalized by the absolute deviance score) was revealed between principals from schools with a response rate of less than 65 % and principals from schools with a response rate of more than 65 %. No significant difference between teaching staff with higher or lower response rate could be shown concerning the total absolute difference ($t = 0.17$; $df = 43$; $p = .99$).

Figure 2: Differences in strain evaluation between principal and teaching staff ($n = 45$)



This analysis of deviance scores does not take into account that the judgments of individual teachers *within* a teaching staff vary considerably. In order to accommodate this vagueness of difference scores, a further categorical strategy of analysis was used. Principals' scores more than a standard deviation above the mean teaching staff score in their own schools were categorized as *overestimations*. If a judg-

ment was more than one standard deviation below the mean teacher evaluation, it was assessed as an *underestimation*.

If the principal estimation was within an area of plus/minus one standard deviation it was assessed as a *match*. According to this criterion in 27 out of 45 cases (60 %), principal evaluation matched with the teacher data. Twelve principals evaluated the strain experience of their teaching staff as less severe than their teachers did; in only six cases, the principals assumed a higher strain experience in their school than the teachers themselves claimed.

4. Discussion

4.1 Summary

Analysis of the aggregated evaluations of 1,643 teachers from 45 teaching staffs indicates that both straining and relieving aspects are perceived in everyday work. In all, neither severe strain nor severe relief is felt in everyday school life.

Principals consider the overall strain experience in their schools, on average, somewhat lower than teachers. This is particularly attributable to the fact that principals evaluate social conditions as more relieving than their staff. Also, they do not expect their staff to feel conference and committee activities to be as straining as they do. In contrast to the overall trend, principals *over*-estimate the strain experienced through activities within the context of actual reforms.

A glance at the correlation of principal and teacher evaluations at individual schools indicates that, overall, accuracy can be optimized. On average, there are 0.68 scale points between the two evaluations, whereby on average, there is a clear underestimation of the strain experienced. Even when considering the internal variation within teaching staffs, about 40 % of the principals' evaluations (18 out of 45) still have to be categorized as a mismatch. Twelve of the principals from these 18 schools underestimate the stress expressed by their staff while six overestimate it. The correlation between ranking orders regarding the severity of strain from both principals and teachers is also low.

4.2 Perception of strain

Before we start with interpreting the results of this study, we want to indicate that the underlying sample does not represent a random sample of the school types involved. In particular, those schools in which the staff only completed a small part of the survey were systematically excluded. It cannot be excluded that a correlation exists regarding individual experiences of strain and participation in the study, so that strain could be underestimated in this sample. Due to the lack of representativeness, it is not possible to infer to the strain experiences in teachers in North

Rhine-Westphalia from this sample. All results concerning mean values of strain of teachers (from the teachers' as well as from the principals' perspective) remain descriptive results for this sample.

A look at the perception of the individual areas identifies two particular facets of work as being straining: conference and committee work and the implementation of reforms and innovations. The negative evaluation concerning conferences is also documented in other studies (Rusteberg, 2004). Teachers express that conferences are frequently unproductive, unrewarding and tedious. Here, particularly the organization of the work rather than the content appears to be the cause of strain. The severe strain experienced through school reforms demonstrated in the present study deserves particular notice. With regards to this finding, it would be interesting in the future to find out how these negative evaluations come about. In addition to the effects of strain caused purely by temporal factors (measurable in working hours), it appears plausible that other psychological factors are also in operation. Reactance reactions to objectives and goals prescribed from above without any participation from those who are affected could play an especially decisive role here.

However, it also should be mentioned that social working conditions were perceived as relieving by most teachers. It appears here that the schools which participated in the study have been successful at creating a positive working environment for their staff. This social resource could reduce the overall level of strain in two ways. Firstly, it could serve as compensation for straining working conditions; secondly, it could contribute to certain objectively existing strains not being perceived as straining by teachers (Rudow, 1994). Regardless of the mode of action, the basic importance of a positive social work environment is to be emphasized – not least from the perspective of a salutogenic approach, which explicitly addresses mental and social well-being and the importance of social resources.

While the data presented here are capable of identifying areas which are associated with high strain experiences, further empirical studies are required to replicate these results with representative samples. Also theoretical considerations and empirical studies are needed to come to conclusions about the actual causal mechanisms in the genesis of perceived strain on teachers. These will then allow for deductions about specific interventions.

Looking at the global score assessing work-related strain in teachers as a whole, we only find a relatively low degree of strain. Neutral and relieving working areas are contrasted by areas that are negative and perceived as straining so that, overall, a neutral value results. Is this to be considered an all-clear? Can it be concluded that teachers in our sample do not perceive their work to be as straining and demanding as is frequently suggested? This conclusion appears somewhat premature for the following methodological reasons. As stated in the introduction, strain was assessed by means of evaluations of generic activity areas and job characteristics. Although the deduction of these particular aspects is oriented on theoretical and empirical knowledge and therefore scientifically based, the question remains

as to whether the surveyed areas are indeed pivotal influences on the experience of strain and whether they completely indicate the potential stressors.

Furthermore, it should be noted that the overall value was calculated as an arithmetic mean value across the different areas. This implies that all areas have the same emphasis on and significance for the experience of strain. In addition, this method of calculation assumes that relieving working conditions have the ability to neutralize negative and straining conditions. Additional effects resulting from the combination of different strain factors were not considered either. Theoretical and empirical analyses of these model assumptions remain to be done.

However, the issue of representativeness plays only if at all a minor role in answering our main research question regarding the accuracy of judgment, because we do not have any theoretically or empirically based assumption that the level of teachers' strain should influence the accuracy of the principals' perception of this strain.

4.3 Principals' accuracy of judgment

As already mentioned in the introduction, the principals are vitally important for the promotion of teacher health, and with it the reduction of strain. A competent implementation of this task, however, requires principals to have an assured ability to accurately perceive the strain experienced by their colleagues. The present findings allow for an initial description of principals' accuracy of judgment.

If one considers the assessment of strain in its entirety within a teaching staff to be a complex and highly-inferential task, it is fundamental to state that many principals' evaluations were fairly successful. Overall, however, they tend to underestimate the strain on their staff. They assess the general social conditions (even) more positively than their teaching staff while conference and committee work is evaluated as being less straining. It is precisely the misevaluation of conferences that is remarkable because this is an occupational activity in which principals can observe their staff directly. Various mechanisms could be the cause of the distortion of judgment here. It may be that teachers conceal their negative perspectives of the situation. It is equally conceivable that principals misinterpret indications of strain because of their particular function and as such the different perspective they have on these committees. Finally, principals are as a general rule themselves highly involved in these situations with regards to content. Such concentration on content makes perception and situational assessment of other parties more difficult. Instead, the evaluation may be primarily too strongly influenced by the principals' personal perceptions. This assumption is compatible with the fact that strain caused by school reform processes, which are initiated externally, is the only area that is systematically overestimated by the principals. It appears plausible that principals themselves experience a high level of strain here and, accordingly, expect their teaching staff to express themselves in a similar manner. This presumption suggests that the evaluation of strain among teaching staff caused by reforms

correlates to the principals' own evaluation of strain caused by regulations of the Ministry of Education ($r = .49$; $p < .001$). In further studies, the evaluation of individual strain in other areas should also be analyzed to examine whether the principals' personal perception of situations is a decisive predictor of suspected strain among teaching staff.

Overall, the findings clearly show the importance of a systematic acquisition of data for the 'diagnosis' of the strain experienced. The task of a health-oriented school design is both too important and too complex for principals to rely on their empathy, their own emotions and perceptions or on chance one-to-one interviews. Therefore, an instrument should be provided that measures teachers strain experiences as valid, reliable and objective as possible. Therefore, based on insights from research into stress and strain (not only in the context of schools), practicable instruments should be developed which will enable principals to examine the strain on their staff in a differentiated manner, thereby broadening and improving their individual judgments. On the basis of corresponding information, problems and requirements for action in the communication process can be worked out, and solutions and interventions can be initiated. Through systematic evaluation, the spiral process of health-oriented school design will be driven forward. However, the starting point of purposeful health promotion must be the accurate evaluation of strain and the necessary action resulting from it. The present study makes it clear that there is a definite need for improvement on the part of principals. The development of a diagnostic apparatus on the one hand, and the investigation of perception distortion by principals on the other, are important and challenging scientific tasks which simultaneously promise high practical benefits for the school context and for the health management which is to be established there.

References

- Blase, J. J., Dedrick, C., & Strathe, M. (1986). Leadership behaviour of school principals in relation to teacher stress, satisfaction, and performance. *Journal of Humanistic Education and Development, 24*, 159–171.
- Borg, M. G., Riding, R. J., & Falzon, J. M. (1991). Stress in teaching: A study of occupational stress and its determinants, job satisfaction and career commitment among primary schoolteachers. *Educational Psychology, 11*, 59–75.
- Burke, R. J., Greenglass, E. R., & Schwarzer, R. (1996). Predicting teacher burnout over time: Effects of work stress, social support, and self-doubts on burnout and its consequences. *Anxiety, Stress and Coping: An International Journal, 9*, 261–275.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Psychology Press.
- DeFrank, R. S., & Stroup, C. A. (1989). Teacher stress and health: Examination of a model. *Journal of Psychosomatic Research, 33*, 99–109.
- Education Act for the Federal State of North Rhine-Westphalia. (2005). *SchulG vom 15.02.2005*. Retrieved from <http://www.schulministerium.nrw.de/BP/Schulrecht/Gesetze/Schulgesetz.pdf>

- Federal Statistical Office of Germany. (2008). *Finanzen und Steuern. Versorgungsempfänger des öffentlichen Dienstes*. (Fachserie 14, Reihe 6.1). Wiesbaden: Federal Statistical Office of Germany.
- German PISA Consortium. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske & Budrich.
- Gieske, M., Harazd, B., & Gerick, J. (2010). Was Lehrkräfte stresst – Das Beanspruchungspotenzial von schulischen Arbeitsbedingungen. *Beruf: Schulleitung*, 1, 16–17.
- Griffith, J., Steptoe, A., & Cropley, M. (1999). An investigation of coping strategies associated with job stress in teachers. *British Journal of Educational Psychology*, 69, 517–531.
- Häbler, H., & Kunz, A. (1985). *Qualität der Arbeit und Verkürzung der Arbeitszeit in Schule und Hochschule*. Munich: IMU-Institute for Media Research.
- Harazd, B., & Gieske, M. (2009). *Schulische Arbeitsbedingungen und ihr Beanspruchungspotenzial*. Poster presented at the 73rd Meeting of the Work Group for Empirical Pedagogical Research (AEPF – Arbeitsgruppe für Empirische Pädagogische Forschung). Bochum: Ruhr-University.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, 59, 297–313.
- Hosenfeld, I., Helmke, A., & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. Prenzel & J. Doll (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (pp. 65–82). Weinheim: Beltz.
- Kramis-Aebischer, K. (1995). *Stress, Belastungen und Belastungsverarbeitung im Lehrerberuf*. Bern: Haupt.
- Kyriacou, C. (2001). Teacher stress: Directions for future research. *Educational Review*, 53, 27–35.
- Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: McGraw Hill.
- Leithwood, K., Menzies, T., Jantzi, D., & Leithwood, J. (1999). Teachers burnout: A critical challenge for leaders of restructuring schools. In R. Vandenberghe & A. M. Huberman (Eds.), *Understanding and preventing teacher burnout – A sourcebook of international research and practice* (pp. 85–114). Cambridge: University Press.
- Matthews, K. A., Cottington, E. M., Talbot, E., Kuller, L. H., & Siegel, J. M. (1987). Stressful work conditions and diastolic blood pressure among blue-collar factory workers. *American Journal of Epidemiology*, 126, 280–291.
- McGrath, J. E. (1981). Stress und Verhalten in Organisationen. In J. R. Nitsch (Ed.), *Stress. Theorien, Untersuchungen, Maßnahmen* (pp. 441–499). Bern: Huber.
- Ministry of School and Further Education of the Federal State of North Rhine-Westphalia. (2006). *Qualitätstableau für die Qualitätsanalyse an Schulen in NRW*. Retrieved from http://www.schulsport-nrw.de/info/o1_schulsportentwicklung/gutersportunterricht/pdf/qualitaetstableau.pdf
- Ministry of School and Further Education of the Federal State of North Rhine-Westphalia. (2007). Das Alter der Lehrkräfte in NRW – Teil 1. *Schulverwaltung NRW*, 4, 127.
- Rudow, B. (1994). *Die Arbeit des Lehrers. Zur Psychologie der Lehrertätigkeit, Lehrbelastung und Lehrergesundheit*. Bern: Huber.
- Rusteberg, D. (2004). *Pilotprojekt bugis – Beteiligungsgestützte und umsetzungsorientierte Gefährdungsbeurteilung an Schulen in Niedersachsen*. Retrieved from http://www.wa.uni-hannover.de/wa/PROJEKT/bugis/Zusfsg_Ergebnisse_bugis.pdf

- Schaarschmidt, U. (Ed.) (2005). *Halbtagsjobber? Psychische Gesundheit im Lehrerberuf – Analyse eines veränderten Zustandes*. Weinheim: Beltz.
- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27–52.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85–95.
- Udris, I., & Frese, M. (1988). Belastungen, Fehlbeanspruchungen und ihre Folgen. In D. Frey, C. Graf Hoyos, & D. Stahlberg (Eds.), *Angewandte Psychologie. Ein Lehrbuch* (pp. 427–447). Munich: PVU.
- van Dick, R. (2006). *Stress und Arbeitszufriedenheit bei Lehrerinnen und Lehrern. Zwischen „Horrorjob“ und Erfüllung*. Marburg: Tectum.
- van Ophuysen, S. (2009). Die Einschätzung sozialer Beziehungen der Schüler nach dem Grundschulübergang durch den Klassenlehrer. *Unterrichtswissenschaft*, 37, 330–346.
- van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz. Eine theoretische und empirische Annäherung. In N. Berkemeyer, W. Bos, H. G. Holtappels, N. McElvany, & R. Schulz-Zander (Eds.), *Jahrbuch der Schulentwicklung* (Vol. 16, pp. 203–234). Weinheim: Juventa.
- Wild, K.-P. (1991). *Identifikation hochbegabter Schüler: Lehrer und Schüler als Datenquellen*. Heidelberg: Asanger.
- WHO – World Health Organization. (1986). *Ottawa Charter for Health Promotion*. Ottawa. (First International Conference on Health Promotion).