

Chapter 6

Taking Composition and Similarity Effects into Account: Theoretical and Methodological Suggestions for Analyses of Nested School Data in School Improvement Research



Kai Schudel and Katharina Maag Merki

6.1 Expanding the Concept of Group Level in School Research

Increasingly, theoretical and empirical studies have shown that the teaching staff plays an important role in school improvement and in fostering student learning, since regulations, guidelines, and the decisions on the system level and on the level of the school management (school leader) have to be re-contextualized by the teaching staff and individual teachers to exert their influence on student learning and student outcomes (Fend, 2005, 2008; Hallinger & Heck, 1998). To deal with such processes, multilevel analysis has proven to be the standard in empirical school research (Luyten & Sammons, 2010). In this contribution, the multilevel approach is expanded to include a theoretical and methodological focus on the double character of group levels in organizations, on composition effects on a group level, and on position effects on an individual level.

Multilevel models allow depiction of hierarchically structured phenomena, such as schools or classes. For example, separate students are gathered in a single classroom, which is often assigned to a specific teacher. Separate teachers, in turn, form a teaching staff and a school, and separate schools are administrated by a school board in a municipality. Finally, schools are part of a geographical entity.

Analysing this nested or clustered structure as a multilevel model is a methodological necessity for two reasons. First, it considers the fact that observations of the same unit are not independent. Thus, it counteracts overestimation of statistical findings, as observations that belong to the same unit on a higher level are interdependent. It also allows determination of the contribution of the different

K. Schudel (✉) · K. Maag Merki
University of Zurich, Zurich, Switzerland
e-mail: kai.schudel@ife.uzh.ch

© The Author(s) 2021

A. Oude Groote Beverborg et al. (eds.), *Concept and Design Developments in School Improvement Research*, Accountability and Educational Improvement, https://doi.org/10.1007/978-3-030-69345-9_6

levels regarding the overall variance of an interesting feature on the lowest level (Luyten & Sammons, 2010). Therefore, differences in student achievement, for example, can be attributed in a more differentiated manner to influences of the separate students, teachers, school management, the school, and possibly also to city districts.

But the way that nested structures are usually considered and calculated by multilevel models indicates a limited understanding of what non-independence of observations within a unit or a group means. This becomes clear by the fact that measures of agreement, such as the intraclass correlation (ICC), is usually used to determine the necessity of a multilevel model. Intraclass correlation (ICC) represents the ratio of the variance between units to the total variance, and it is interpreted as a measurement of agreement or similarity among observations within a unit (LeBreton & Senter, 2007). Therefore, when non-independence is conceived of only as the presence of a significant ICC value, the non-independence is simply defined by an over-proportional similarity of observations within a unit. But non-independence can mean more than converging observations, such as, for example, same shared attitudes among teachers or the same teaching staff. Non-independence in nested structures can be defined more generally by simply acknowledging that observations are influenced by the unit that they are in, and thus, by the shared context, and the unit's influence can manifest itself in various forms. For teachers on a teaching staff, for example, the shared unit does not have to lead to shared attitudes. The same shared unit can also result in different attitudes because the teaching staff serves as an umbrella under which teachers have to interact. In this sense, non-independence means that every teacher refers to the other teachers within the same teaching staff. Thus, each teaching staff can be described by a specific composition and pattern that are a result of non-independence of the teachers.

This problem of too simplified group-level conceptions and non-independence has also been criticized in research on small groups and in organizational research by Kozlowski and Klein (2000). They also point out that research often simply aggregates lower-level individual characteristics to the next higher group level by averaging, without considering that groups can also be described by the specific composition of the individual characteristics. They suggest that groups and, thus, every higher level in nested data can be described by *global properties*, *shared properties*, and *configural properties*. We can adopt these aspects in our criticism of school research above. Global properties are located at the group level, or the higher level, respectively; they manifest only on that level, and their measurement does not depend on lower-level characteristics and are thus non-controversial. Therefore, global properties of a group serve as a shared context for lower level individuals. Furthermore, because they serve as a context for the individuals on lower level, global properties initiate a top-down process (Kozlowski, 2012). Collective characteristics of the lower level, which describe how similar or dissimilar group members are, can be generally described by *group composition* (Kozlowski, 2012; Lau & Murnighan, 1998; Mathieu, Maynard, Rapp, & Gilson, 2008; Schudel, 2012). According to Kozlowski and Klein (2000), the composition of a group can be described by shared properties or by configural properties. Shared properties are

those characteristics of individuals that converge within the group and represent the homogeneity thereof. Configural properties are those characteristics of individuals that diverge within the group and represent the heterogeneity of a group.

In the case of school research, the neglect of group composition may be connected to the *double character* that group levels in school environment usually possess. The entities on a higher level – such as schools or classrooms – can be described by either separate characteristics on that higher level – the global properties – or by collective characteristics on a lower level – the group composition. Global properties can be an area of responsibility of a single individual on the higher level or a shared higher-level context. However, collective characteristics on a group level can only be described by the interplay of multiple individuals on the lower subordinate level. They emerge from the lower level by interaction but manifest themselves at the group level; thus, group composition refers to the fact that what develops in a group is more than just the simple sum of the individuals (Kozłowski & Klein, 2000). Therefore, the information about the global properties of a group can be obtained from that group level, and the information about group composition can only be gathered from the multiple lower level entities. For instance, if we are interested in the school level, we can describe and measure the global properties by separate characteristics of the responsible school principal or of the school, such as leadership quality and budget. But we can also describe and measure the composition of the school by collective characteristics of the cluster of teachers working at the school, the shared and configural properties of the teaching staff, such as shared beliefs of the teachers, but also as diverging subjective perspectives. The same holds true for the classroom level: We can describe and measure the global properties by separate characteristics of the responsible class teacher or of the classroom infrastructure, such as teaching quality and the number of computers available. We can also describe and measure the classroom composition by collective characteristics of the cluster of students that form a class, e.g. the average school achievement of the students as a shared property, when we assume that students in a class tend to have a similar learning progress – or e.g. different educational family backgrounds as a configural property.

In conclusion, although multilevel models in school research acknowledge that a group level always constitutes a combination of entities of a lower level (e.g. teaching staff as an association of teachers), the underlying assumption usually is that the shared group context leads to homogeneous entities. Therefore, research often focuses solely on shared properties, which is represented by the calculation of a group mean. However, the explanations above show that non-independence and shared group context do not preclude the possibility that the lower-level entities or individuals are different. Therefore, multilevel models in school research have to consider the double character of groups, consisting of global group properties emerging from the group level, and group composition emerging from the individual lower level. Further, they have to consider the possibility of both shared properties and configural properties of group compositions.

Disentangling those two characteristics of a group or a higher level entity is also crucial because it allows us to depict the re-contextualization processes in the school

environment (Fend, 2005, 2008). If we separated global properties from group composition, we could make it visible that global properties – such as a responsible person or an existing infrastructure – serve as an opportunity and that individuals on the lower level make use of that opportunity by their specific group composition. Kozlowski (2012) analogously observes that a group is finally the result of top-down effects of global properties and bottom-up effects emerging from the group composition. That what we measure on a specific unit level, therefore, is mostly a result of the interactions between a responsible separate person, or a shared context characteristic, and a subordinate collective as shown in Fig. 6.1.

As composition and configural properties in particular are often missing in research, we can assume that research reduces unit levels to areas of responsibility rather than also take their collective character of associations into account. Therefore, contrary to the theoretically acknowledged fact that diversity of the teaching staff has an influence on school improvement processes, research has placed too little emphasis on the compositional characteristics and composition effects of the teaching staff in study designs and analyses.

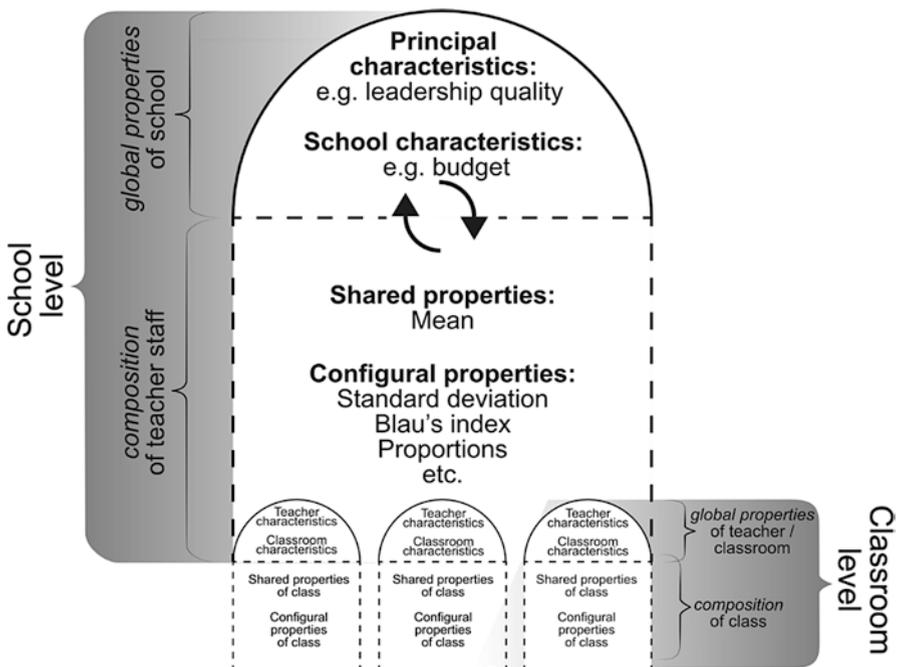


Fig. 6.1 Double character of group levels in school research
 Group levels can be described by separate global properties (semi-circles) and by collective composition (dashed rectangles). Group compositions emerge from subordinate lower level entities and can be described by shared properties and by configural properties. A group is a product of top-down effects of global properties and bottom-up effects of group composition

At class level, the well-known ‘little-fish-big-pond effect’ can be taken as an example: A student’s self-concept is affected not only by his or her own achievements, but also by the aggregated average performance index of the classroom (the entity one level above the student). Accordingly, the school class acts as a frame of reference, through social comparison, for students’ self-concepts (Marsh et al., 2008). This is a phenomenon at the classroom level, and it has also been understood as a composition effect.

Further, pertaining to the level of the teachers, the literature on school improvement capacity or professional learning communities points to the importance of group composition. Mitchell and Sackney (2000), for example, emphasize the relevance of interpersonal capacities to learning communities. This relevance becomes apparent in shared properties, such as shared norms, expectations, and knowledge, or in communication patterns, among other things. For group climate to be effective, each group member’s contributions should be explicitly acknowledged. As a consequence, Mitchell and Sackney (2000) also observed problems in schools with high configural properties, thus, with group compositions, in which dominant excluding subgroups were formed that isolated and marginalized other members. Also, Louis, Marks, and Kruse (1996) showed that diverse subgroups within the teaching staff can have negative effects on the successful achievement of joint objectives. They assume that subgroups can emerge particularly in large schools, alongside discipline demarcations. However, despite the relevance of the composition and structure of teaching staff, there are (still) no studies examining these composition effects differentially.

Based on diversity research, we will first elaborate on how composition can be theorized in school improvement research, particularly at the teaching staff level. In a second step, the Group Actor-Partner Interdependence Model (GAPIM) approach is introduced as a methodological tool. The GAPIM allows analysis of composition effects on the individual level and takes the particular position of the teachers on staff into consideration. We then apply the model to an existing data set (Maag Merki, 2012) as an example.¹ We will illustrate the analysis of the main effects and composition effects of the teaching staff and positioning effects of the separate teachers on the teaching staff regarding the effects of teachers’ individual and collective self-efficacy on teachers’ individual job satisfaction. Since in the existing study, teachers at 37 secondary schools completed a standardized survey on various aspects, the data set is suitable to discuss strengths and weaknesses of the GAPIM for school improvement research.

¹Originally, Maag Merki (2012) analyzed the effects of the implementation of state-wide exit examinations on school, teachers, and students in 37 German upper secondary schools (ISCED 3a). The present contribution, however, does not focus on the analyses of the effects of the implementation of state-wide exit examinations.

6.2 Composition Effect as Diversity Typologies

As mentioned above, the composition of a group can be described by converging or diverging characteristics represented by shared and configural properties. In order to conceptualize different types of shared and configural properties, approaches from diversity research and particularly the typology of Harrison and Klein (2007) are useful (Schudel, 2012).

Diversity of teams is of great importance in the concept of learning communities and distributed leadership (Hargreaves & Shirley, 2009; Mitchell & Sackney, 2000; Stoll, 2009). But diversity can have diverging consequences. It can lead to lower levels of communication through social categorization processes, but [at the same time] it can lead to higher levels of problem solving when diversity reflects a variety of different qualities (Van Knippenberg, de Dreu, & Homan, 2004; Van Knippenberg & Schippers, 2006). This twofold character of diversity is a central issue in research on small groups and is discussed theoretically from an interference-oriented perspective and a resource-oriented perspective (Schudel, 2012). In the context of school improvement, Mitchell and Sackney (2000) point out that diversity endangers a teaching staff, if it leads to the formation of subgroups and, in doing so, undermines shared norms and cooperation. In contrast, the potential of diversity is expressed in the demand “to make a cultural transformation so as to embrace diversity rather than to demand homogeneity” (Mitchell & Sackney, 2000, p. 14). A more differentiated theoretical account of diversity is needed in order to account for the composition effects of teams.

Harrison and Klein (2007) differentiated three types of diversity: *separation*, *variety*, and *disparity*. This differentiation provides a basis for both the interference-oriented perspective and the resource-oriented perspective. With *separation*, diversity can be described as a measure for the formation of subgroups. It is based on similarities between group members regarding a distinct feature, a position or opinion, quantified along a continuum. Consequently, teachers can be compared with each other, for example regarding their tenure – i.e. their position along the continuous attribute tenure. Separation describes the level of similarity between group members. This level is expressed statistically through the standard variation of the feature on the group level. Therefore, a teaching staff exhibits a high level of separation, if the teachers hold positions on both extreme poles of the specific feature’s continuum, such as when half of the teachers have only recently been employed at the school while the other half have been working there for a long time. There is a moderate degree of separation when the teachers are distributed evenly over the continuum of the feature. There is a small degree of separation when all teachers hold the same position on the continuum of the feature, such as when they all have been employed at the school for an equally long time. Since separation is a symmetrical similarity measure, it would be irrelevant at a low level of separation, if all teachers exhibited a long or a short term of employment. Relevant would only be that they exhibited a similarly long or similarly short term of employment. Therefore, separation constitutes a conceptualization in accordance with the practically

relevant potential of subgroup formation within a teaching staff. From an interference-oriented perspective, high separation would have negative consequences for communication and interaction.

The second type of diversity, following Harrison and Klein (2007), is *variety*. The term variety describes the presence of different resources and qualities within a group. It is based on different features of group members that are not quantitatively comparable on a continuum but are of different qualities. For example, teachers are able to form a more or less diverse and heterogeneous teaching staff regarding their subject(s), function, or discipline. Therefore, variety describes the heterogeneity of categorically different features or qualities. Statistically, this is expressed in Blau's index (1977), describing the number of different categories available within a group. Therefore, the teaching staff possesses the highest variety, if all members of the teaching staff teach a different subject, for example. There would be minimal variety in this respect, if all teachers taught the same subject, or, in other words, if the school was highly specialized. Variety is thus operationalized as the different qualitative backgrounds of the teaching staff. It reflects the presence of different kinds of knowledge and abilities in the sense of informational diversity. From a resource-oriented perspective, high variety could therefore be beneficial for problem-solving in community learning (Jehn, Northcraft, & Neale, 1999). Yet, from an interference-oriented perspective, high variety could also describe potential difficulties for divided norms and values and commitment in big and fully differentiated schools (Louis et al., 1996).

Finally, as a third type of diversity, *disparity* means the distribution of hierarchically structured resources within a group. It is based on the distribution of certain normatively desired or valuable features within a group – such as power, wealth, status, or privileges – that are understood as scarce resources. Disparity is, therefore, an asymmetrical measure. It makes a difference whether a minority or a majority holds most of the resources. For example, teaching staffs can differ in how competencies and decisional power are equally distributed among the teachers. Statistically, disparity is expressed in the proportional relation between group members and resource allocation. The teaching staff exhibits a high level of disparity, if, for example, a minority of teachers possess the most – or an unproportioned amount of – decisional power. A lower level of disparity prevails, if the teaching staff has a flat hierarchy, and all teachers have a similar amount of decision-making authority. Disparity is thus able to describe, for example, how much say the teachers have in important decisions and how strongly they are included/involved in the development of changes. Disparity can offer an important indicator of the distributed leadership status (Stoll, 2009).

The three diversity types describe the composition of groups. Instead of reducing the teaching staff to its shared properties and solely considering its group means, school improvement research has to take the multi-faceted composition of the teaching staff into account. Furthermore, Harrison and Klein's (2007) diversity typology not only reveals additional important *descriptive* information about characteristics of shared and configural properties of the teaching staff, but can also be used in causal analyses. The composition measures of the teaching staff can be modelled as

results of antecedent processes. Good school leadership, for example, can result in a teaching staff with low separation, high variety, and low disparity. Or, alternatively, the composition measures of the teaching staff can be modelled as causes of the outcomes of schools, teaching staffs, and separate teachers. For example, from an interference-oriented perspective, high separation of a teaching staff can result in low performance of the school, in low cooperation within the teaching staff, and in low job satisfaction in separate teachers. As a result, these measures introduce new insights into school development research regarding how the teaching staff is structured, what causes this structure, and to what extent the structure has an influence on teacher outcomes, the development of curricula, or the learning curve of students.

6.3 Positioning Effect

Now, if group compositions of this kind are to be examined as predictors of dependent variables on a subordinate individual level, the three diversity types by Harrison and Klein (2007), presented above, have theoretical and methodological shortcomings. Further considerations are necessary that incorporate the individual level.

Diversity, conceptualized on only the group level, abstracts from the *definite position* of the single individual within the group. However, if group composition is taken as a predictor of effects on the individual level, this definite position of the individual within the group composition will not be ignored. Accordingly, group composition signifies different things, depending on the position of a person within this diversity. Naturally, this is most evident in the asymmetrical group composition of disparity. For example, depending on where teachers are within a group characterised by a high level of disparity, they are in possession of resources or not. But also regarding symmetrical measures, such as separation and variety, there are differences in teachers' positions within the compositions of their groups. For example, a group might exhibit a low level of separation or variety. Yet, if a single teacher deviated from such an otherwise homogeneous group, that person could perceive their individual position as isolated. A moderate separation of the teaching staff regarding tenure can have different effects for those teachers that exhibit average tenure (and, thus, are positioned along the continuum in the middle) as compared to newly employed teachers and the most senior teachers (and, thus, those positioned at one of the extreme poles).

Kenny and Garcia (2012) describe this definite position within a group by means of *similarity* relations between the individual and the *rest* of the group. They emphasize that “the key conceptual and psychological contrast in groups is between self and others and not between self and group” (Kenny & Garcia, 2012, p. 471). Indeed, people primarily perceive themselves not as contrary to a group average but rather as opposites to the rest of a group. Consequently, for specific teachers, the homogeneity and heterogeneity of their group always take the form of similarities between themselves and the others in their group into account. Kenny and Garcia (2012) proposed to model such an inclusion of separate positions within a group and their

similarities with the rest of their group using the Group Actor-Partner Interdependence Model (GAPIM), which will be outlined in the following section.

6.4 Modelling Position Effects

Using the GAPIM, the individual value of an interesting feature of a group member is conceived as the result of four different terms or predictors: *actor effect* X , *others' effect* X' , *actor similarity* I , and *others' similarity* I' . A group member is defined as the actor and the rest of the group as the others. The actor effect designates the influence of an independent variable of a group member on its dependent variable, for example the influence of self-efficacy on one's own level of satisfaction. The others' effect then designates the influence of the average of the same independent variable of the others on the dependent variable of the actor. With these two main effects, Kenny, Mannetti, Pierro, Livi, and Kashy (2002) revised the classical multilevel analysis. Group effect, or influence of the group level, is not included as usual in the analysis as total group value; only the average value of the others is included in the GAPIM. In doing so, the influence of the actor is partialized out of the group value.

In addition to the two main effects, *actor effect* and *others' effect*, there are two similarity effects for the study of composition effects. These are based on actor similarity, which models the similarity between the actor and every single other group member regarding an independent variable. Others' similarity models how similar the others are to each other. These similarity terms represent values for the respective position of the actor within the group regarding the independent variable. In addition, these values can now be entered into the analysis as well, whereby the influence of the similarity between actor and others, and among the others, on the dependent variable of the actor can be calculated. In this way, a group composition *from the perspective* of each group member can be modelled. Hence, a value on the individual level is predicted on the basis of two main effects and two similarity effects. If the level of actor similarity is high, the actor is in a numerically more dominant subgroup or in a more homogeneous overall group; if it is low, the actor is isolated from the rest of the group, or at least from every single other in the group. If the level of others' similarity is high, the rest of the group is homogeneous and forms a dominant subgroup, or a homogeneous overall group together with the actor. For an extremely isolated teacher, there is low actor similarity and high others' similarity; thus, the teacher is confronted with a homogeneous, numerically dominant subgroup, of which he or she is not a member. In contrast, when there is high actor similarity and high others' similarity, then the teacher is part of a homogeneous subgroup.

According to Kenny and Garcia (2012), an individual value of a dependent variable (Y_{ik}) consists computationally of a constant (b_{0k}), the four outlined effects ($b_1X_{ik}; b_2X'_{ik}; b_3I_{ik}; b_4I'_{ik}$), and an error term (e_{ik}):

$$Y_{ik} = b_{0k} + b_1X_{ik} + b_2X'_{ik} + b_3I_{ik} + b_4I'_{ik} + e_{ik}$$

Note that $b_2X'_{ik}$, b_3I_{ik} and $b_4I'_{ik}$ constitute effects that relate to the others in the group or to the teacher's relation to the others in the group. Therefore, they are included computationally on the individual level in the present analysis.

In addition, to examine socio-psychological group theories, the four terms can be coded in such a way that different group compositions can be estimated by contrasts, fixations, and equations and compared with each other via model fit (Kenny & Garcia, 2012). With these submodels, it can be determined to which features group members react more sensitively regarding composition effects in general. Accordingly, the two main effects can be analysed in a *Main Effects Model*; the actor effects can be solely analysed in the *Actor Only Model*; and the others' effects can be solely analysed in an *Others Only Model*. In the *Group Model*, actor and others' effects are equated with each other, whereby this model represents the classical multilevel model. Finally, in the *Main Effects Contrast Model*, actor and others' effects are contrasted.

The inclusion of similarity effects thus allows for more differentiated modelling possibilities than have been available up to now. In a *Person-Fit Model*, where the suitability of the separate group member regarding the rest of the group matters, the inclusion of actor similarity in addition to the main effects leads to the best model fit. In a *Diversity Model*, where diversity in the whole group matters, the inclusion of both similarity effects in addition to the main effects leads to the best model fit. In a *Complete Contrast Model*, where the contrast between actor similarity and others' similarity matters, the complementary coding of the similarity effects in addition to the main effects leads to the best model fit. Finally, if all four terms are included without constraints, we refer simply to a *Complete Model*.

6.5 Present Study: The Relation Between the Influence of Composition and Similarity Effects on Job Satisfaction

The advantages of the GAPIM over a conventional multilevel analysis will be illustrated by means of an example from school research. Based on a data set from a study on the effects of the introduction of state-wide exit examinations on schools, teachers, and students (ISCED 3a) (Maag Merki, 2012), we analyse how motivational characteristics of teachers – individual teacher self-efficacy (ITE) and perceived collective teacher self-efficacy (CTE) – affect job satisfaction. With this, we focus on an example that deals with teachers at the individual level and with the teaching staff of the school at the group level. We calculate the influences of the main effect on the group level (group mean), the composition effect on the group level (standard deviation), the main effects on the individual level (actor effect and others' effect), and the position effects on the individual level (actor similarity and others' similarity) on individual job satisfaction.

The two self-efficacy variables qualify for the GAPIM for two reasons: First, in accordance with 'big-fish-little-pond effect' research (Marsh et al., 2008), it can be

assumed that motivational characteristics are especially sensitive to composition and positioning effects because comparison processes with the 'others' are crucial. Second, the two self-efficacy variables share a conceptual similarity, albeit on different levels (individual and group level).

The two concepts, ITE and CTE, refer to Banduras' (1997) concept of self-efficacy. They both describe the individual's perception of being able to master future challenges (Schmitz & Schwarzer, 2002). However, ITE describes the perceived abilities and potentials of the separate teachers, whereas CTE describes the teaching staff's collective self-efficacy, which is perceived and assessed on an individual level as well (Goddard, Hoy, & Hoy, 2000; Schwarzer & Jerusalem, 2002). According to Schwarzer and Jerusalem (2002), CTE consists of meta-individual beliefs of the teaching staff concerning being able to manage future events in a positive manner as a team. ITE and CTE correlate with each other, but they can be described as independent constructs because of their only moderately high level of correlation (Schmitz & Schwarzer, 2002). The question arises here as to what extent CTE really represents meta-individual beliefs or whether it only represents ITE at its own level (Schwarzer & Schmitz, 1999; Skaalvik & Skaalvik, 2007).

According to group main, group composition, and individual main and positioning effects explained above, there are three ways that ITE and CTE can have an effect on job satisfaction.

First, self-efficacy beliefs generally exhibit a positive correlation with job satisfaction. Positive correlations have been found regarding general self-efficacy (Judge & Bono, 2001), individual teacher self-efficacy (ITE) (Caprara, Barbaranelli, Borgogni, & Steca, 2003; Klassen, Usher, & Bong, 2010), and collective teacher self-efficacy (CTE) (Caprara et al., 2003; Klassen et al., 2010; Skaalvik & Skaalvik, 2007). Therefore, we expect to find direct main effects of ITE and CTE – on both the individual and group level – on individual job satisfaction. Teachers with high ITE and teachers, who perceived high CTE, should have higher individual job satisfaction. And teaching staffs where teachers report on average higher ITE and CTE should lead to higher individual job satisfaction of the teachers.

Second, we also expect composition effects of ITE and CTE on individual job satisfaction. Various studies show that the teachers' perceptions of their own coping resources or the coping resources of their team can vary within a team (e.g. Moolenaar, Slegers, & Daly, 2012; Schmitz & Schwarzer, 2002). Further, schools differ in their composition of teachers regarding ITE (Schwarzer & Schmitz, 1999). If some teachers on the teaching staff report low levels of ITE and CTE, while other teachers show high levels, then this variation could lead to high levels of separation. From an interference-oriented perspective, this could have a negative effect on individual job satisfaction. Separation of ITE can indicate an actual lack of collective problem-solving processes in the teaching staff, and it should therefore be congruent with the perception of low CTE. In addition, separation of CTE indicates not only that there is a lack of collective problem-solving processes, but also that teachers experience their same teaching staff differently. In this case, some teachers believe in their collective ability to master future problems, while other teachers do not. The separation of CTE indicates disagreement on the way of looking at a

problem. Therefore, teachers on teaching staffs with high separation of ITE and CTE could have lower job satisfaction than their counterparts on teaching staffs with homogeneous ITE and CTE reports.

Third, in addition to individual main effects, we expect to find positioning effects of ITE and CTE on the individual level on individual job satisfaction. The fact of being isolated on a teaching staff could decrease individual job satisfaction. This is obvious for teachers with low ITE on a teaching staff with others having high ITE. However, in the opposite case, too – for teachers with high ITE on a teaching staff with others having low ITE – isolation can have negative effects on individual job satisfaction. Sharing the same fate of low ITE can lead to similar perspectives and collective support and can help build trust and ties. Being barred from such a collective support can harm individual job satisfaction. The same holds true for CTE. But additionally, CTE refers to an individual's perception of a collective characteristic. Therefore, when a teacher's perception of CTE differs strongly from the others' perceptions, it can be assumed that this teacher does not share all collective processes of the teaching staff. Referring to CTE, isolation can thus indicate objective isolation within the teaching staff and can be detrimental to individual job satisfaction. Therefore, in terms of the GAPIM, the others' similarity of ITE and CTE should have a negative effect on job satisfaction, and the actor's similarity of ITE and CTE should have a positive effect thereon.

6.6 Methods

6.6.1 Sample

The study took place from 2007 to 2011 in the two German states of Bremen and Hesse, which introduced state-wide exit examinations at the end of secondary school (ISCED 3sa). Standardized surveys were conducted in 2007, 2008, 2009, and 2011 (Maag Merki, 2016). In total, 37 secondary schools participated, and surveys were administered to teachers and students. In Bremen, all but one secondary school took part in the surveys (19 schools). In Hesse, the schools were chosen based on crucial context factors (e.g. region, urban–rural, profile of the school). The current study used the teacher data from 2008, which was the first year in which the teachers in both states had to deal with state-wide exit examinations.² A sufficiently large school sample ($N = 37$) and teacher samples (total $N = 1526$, $N_{\text{Bremen}} = 577$, $N_{\text{Hesse}} = 949$) were available to be used for the multilevel analyses. The response rate was sufficient, at 59%. The composition of the sample can be regarded as being representative for both Hesse and Bremen regarding teacher gender and amount (hours) of teaching activity. Young teachers were somewhat over-represented and

²As mentioned above, the analyses of the effects of the implementation of state-wide exit examinations are not the focus of this paper.

teachers older than 50 slightly under-represented. Further descriptive statistics are available in Merki and Oerke (2012).

6.6.2 Measurement Instruments

ITE was collected using a scale by Schwarzer, Schmitz, and Daytner (1999) with six items; the scale exhibited a range of 1 to 4 ($\alpha = .74$; $M = 2.84$; $SD = 0.44$). An example item is: “Even if I get disrupted while teaching, I am confident that I can maintain my composure.” The response scale ranged from 1 = not at all true, 2 = barely true, 3 = moderately true, to 4 = exactly true. Since this scale is skewed, it was transformed into an ordinal variable with four categories.

CTE was measured with five items that exhibited a range of 1 to 4 ($\alpha = .76$; $M = 2.54$; $SD = 0.51$) (Halbheer, Kunz, & Maag Merki, 2005; Schwarzer & Jerusalem, 1999). An example item is: “We as teachers are able to deal with ‘difficult’ students because we have the same pedagogical objectives.” The response scale ranged from 1 = not at all true, 2 = barely true, 3 = moderately true, to 4 = exactly true.

Job satisfaction was assessed with six items that exhibited a range of 1 to 4 ($\alpha = .80$; $M = 1.88$; $SD = 0.51$) (Halbheer et al., 2005). The scale entered the study with z-standardization. An example item on the job satisfaction scale is: “I am enjoying my job.” The response scale ranged from 1 = not at all true, 2 = barely true, 3 = moderately true, to 4 = exactly true.

6.6.3 Analysis Strategies

The different theoretical and methodological approaches presented above that consider group characteristics in nested data were compared. For this, we first calculated the measure that is usually considered a requirement for a conventional multilevel analysis, the intraclass correlation (ICC). As described above, ICC states how much of the total variability comes from the variability between teaching staffs and from the variability within teaching staffs. Thus, ICC refers to a limited understanding of non-independence of teacher consensus within a teaching staff. A significant ICC size – tested with the Wald-Z – would then indicate that teachers within a teaching staff are over-proportionally similar. However, a non-significant ICC size would indicate a lack of convergence of teachers and would be interpreted as independence of teachers within a teaching staff. In this case, referring to the conventional procedure, the assumption of nested data would be withdrawn, and there would be no necessity for a multilevel analysis.

Second, we calculated a multilevel analysis to examine, if there was a *main group level effect* of the two self-efficacy variables on the teaching staff level to job satisfaction on the individual level. For this purpose, the group means of ITE

($M = 2.840$; $SD = 0.0949$) and CTE ($M = 2.520$; $SD = 0.1640$) on the teaching staff level were calculated as predictors of job satisfaction on the individual level.

In a third step, we examined if there was a *composition effect* of the two self-efficacy variables on the teaching staff level to job satisfaction on the individual level. In this case, we operationalized composition as separation within the teaching staffs and thus as standard deviation. For this purpose, the standard deviations of ITE ($M = 0.434$; $SD = 0.0651$) and CTE ($M = 0.4813$; $SD = 0.0912$) were calculated on the teaching staff level as predictors of job satisfaction on the separate teacher level.

In a fourth step, we examined *main and similarity individual level effects* on the separate teacher level using the GAPIM. For this purpose, we used Kenny and Garcia's macro for SPSS (Kenny & Garcia, 2012). It is based on the linear mixed model in SPSS. The advantage of the macro is that it automatically calculates main and similarity terms and compares the different submodels with each other according to the fit index SABIC (Sample-size Adjusted Bayesian Information Criterion). In addition, we calculated Chi^2 difference tests to estimate whether some differences between the model fit of submodels were significant; Chi^2 difference tests were based on the log-likelihood values. To calculate the similarity terms, continuous and categorical predictors have to be transformed in such a manner that the lowest value is -1 and the highest value 1 .

For samples in the field, however, the problem of multi-collinearity arises. The main effects tend to covary with the similarity effects regarding skewed predictors. For example, if a sample consists of only a few teachers that scored low on individual self-efficacy, it is more likely that these teachers differ from the other members of the teaching staff, i.e., that the similarity term I is smaller. To counter this confound, the skewed continuous predictor ITE is recoded to an ordinal scale. The continuous variable is divided into quartiles; the new ordinal variable thus consists of four categories with equal amount of cases.

To show the benefits of using the GAPIM, the Actor Only Model is reported with only the main actor effect X . It corresponded to a multilevel model with a predictor variable on the individual level. The Main Effects Model followed by adding the main others effect X' , which describes the average predictor effect of the rest of the teaching staff. In this context, the GAPIM differs from the classical multilevel model because the predictor variable was not included in the analysis on the group level (as group average) but entered the analysis with X' as a variable on the individual level. With the Complete Model, finally, the two similarity terms actor similarity I and others' similarity I' were added, which constitute the specific nature of GAPIM.

6.7 Results

6.7.1 Analysis of Variance

In a first step, we analysed to what extent a multilevel model that follows common criteria is necessary at all regarding the dependent variable job satisfaction. A fully unconditional, or no predictors, model resulted in an insignificant group level variability of 0.01243 with a Wald-Z of 1.540 ($p = .124$) and an intraclass correlation of $ICC = 0.01243$. According to Heck, Thomas, and Tabata (2010), the percentage of variability of the dependent variable that is attributed to the group level is too small to be acknowledged with an ICC value below 0.05.

According to common criteria, a multilevel analysis would be refrained from because it is to be assumed that only a small part of the total variability of job satisfaction is to be attributed to differences between the teaching staffs. As has been argued, this point of view reduces non-independence in nested data to homogeneity within a unit and ignores that non-independence can also be described by specific compositions within units. Refraining from carrying out a multilevel analysis, at this point, could lead to missing information about composition and positioning effects.

6.7.2 Main and Composition Effects

In a second and third step, we analysed the main and composition effects on the teaching staff level on individual job satisfaction. In the linear mixed regression model with group mean of ITE (main effect) and standard deviation of ITE (composition effect) as group level predictors, job satisfaction was predicted only by the group mean, with $B = 0.755$ ($p = .000$). The standard deviation of ITE had no significant effect on job satisfaction ($B = -0.026$; $p = .957$).

The result for CTE was the same: Job satisfaction was predicted by the group mean of CTE (main effect) ($B = 1.151$; $p = .000$). The standard deviation of CTE (composition effect) had no significant effect on job satisfaction ($B = -0.197$; $p = .725$).

Consequently, there are only main and but no composition effects in classical multilevel analyses with predictors on the group level. Teaching staffs with high ITE and CTE levels on average, indeed, showed higher levels of individual job satisfaction. The level of separation between the teachers regarding these variables, however, had no influence on individual job satisfaction.

6.7.3 Main and Similarity Effects with GAPIM and Multilevel Analysis

In a fourth step, we analysed main effects and similarity effects on the individual level on individual job satisfaction.

6.7.3.1 Individual Teacher Self-Efficacy as Predictor

Table 6.1 lists all submodels – the Actor Only Model, the Main Effects Model, and the Complete Model. The Actor Only Model showed that individual job satisfaction was predicted by ITE with $B = .714$ ($p = .000$), and it had a multiple correlation of R^2 of .528. For the Main Effects Model, we included the X' term, i.e. the average ITE of the rest of the teaching staff. But X' had no significant effect, with $B = 0.18$ ($p = .888$). For the Complete Model, we finally included the similarity terms I , i.e. the similarity of the actor compared to the other members of the teaching staff, and I' , i.e. the similarity of the other members of the teaching staff among themselves regarding ITE. The Complete Model showed that ITE still had a positive main effect on the individual level of job satisfaction, with $B = .697$ ($p = .000$). The X' term remained insignificant, with $B = .078$ ($p = .616$), and the I term was insignificant as well, with $B = .210$ ($p = .276$). The I' term had a marginally significant effect, with $B = -1.521$ ($p = .056$), however. This means a teacher’s job satisfaction was the lower, the more the other teachers agreed in their ITE reports. Whenever the other teachers were divided in their ITE reports, then the teacher’s job satisfaction increased. This can be quantified in an example of a teacher on a teaching staff with eleven other teachers: A teacher reported a lower job satisfaction of 1.651 standard deviations while all other teachers reported the same ITE as opposed to when six other teachers reported the lowest ITE and five teachers the highest.

With a lower SABIC of 3656.934 ($R^2 = .529$), the model fit of the Complete Model indeed exceeded the model fit of the Actor Only Model (SABIC = 3660.328,

Table 6.1 Effect coefficient estimations and model fits of ITE on job satisfaction

Model	Main effects		Similarity effects		Model fit	
	X	X'	I	I'	SABIC ^b	R ²
Empty	– ^a	– ^a	– ^a	– ^a	4134.71	.000
Actor only	0.714***	– ^a	– ^a	– ^a	3660.33	.528
Main effects	0.713***	0.018	– ^a	– ^a	3660.79	.528
Complete	0.697***	0.072	0.210	–1.513+	3656.93	.529

Note. X = Actors individual teacher self-efficacy; X' = Others' individual teacher self-efficacy; I = Actor similarity; I' = Others' similarity; SABIC = Sample-size adjusted Bayesian information criterion

+ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

^aFixed to zero

^bSmaller SABIC means a better fitting model

$R^2 = .528$). But the improvement in the model fit was not significant ($Chi^2 = 4.851$; $df = 3$; $p = .183$). However, our primary interest was not in the best fitting model, but in showing that by using the GAPIM, we are able to obtain additional information about positioning effects. In this case, we found that a teacher's job satisfaction was not only positively influenced by its ITE, but was also (in tendency) negatively influenced by the similarity of the rest of the teachers on staff regarding their ITE.

6.7.3.2 Collective Teacher Self-Efficacy as Predictor

Table 6.2 also lists all submodels – the Actor Only Model, the Main Effects Model, and the Complete Model. The Actor Only Model showed that the individual level of job satisfaction was predicted by CTE with $B = 1.356$ ($p = .000$) and had a multiple correlation of R^2 of $.457$. In the Main Effects Model, the additional X' term had no significant effect, with $B = -.180$ ($p = .536$). The Complete Model, finally, showed that CTE still had a positive main effect on the individual level of job satisfaction, with $B = 1.322$ ($p = .000$). The X' term remained insignificant, with $B = 0.115$ ($p = .776$). The I term, i.e. the similarity of the actor to the other members of the teaching staff, was significant, with $B = 1.627$ ($p = .031$), and the I' term was insignificant, with $B = -3.919$ ($p = .128$). This means that a teacher's job satisfaction was the higher, the more similar his or her CTE was to that of the other teachers. This can be quantified: A teacher reported a higher job satisfaction of 3.255 standard deviations, if he or she reported exactly the same CTE as the other teachers on staff than if he or she reported the most divergent CTE compared to other teachers on staff.

With a lower SABIC of 3752.214 ($R^2 = .459$), the model fit of the Complete Model indeed exceeded the model fit of the Actor Only Model (SABIC = 3757.594, $R^2 = .457$), although the improvement in the model fit was only nearly significant ($Chi^2 = 6.837$; $df = 3$; $p = .077$). However, this does not lower the importance of the result that teachers' job satisfaction was positively influenced not only by its CTE, but also by the fact how similar he or she perceived CTE compared to the other teachers on staff.

Table 6.2 Effect coefficient estimations and model fits of CTE on job satisfaction

Model	Main effects		Similarity effects		Model fit	
	X	X'	I	I'	SABIC ^a	R ²
Empty	– ^b	– ^b	– ^b	– ^b	4094.59	.000
Actor only	1.356***	– ^b	– ^b	– ^b	3757.59	.457
Main effects	1.362***	–0.180	– ^b	– ^b	3757.68	.457
Complete	1.332***	0.115	1.627*	–3.919	3752.21	.459

Note. X = Actors individual teacher self-efficacy; X' = Others' individual teacher self-efficacy; I = Actor similarity; I' = Others' similarity; SABIC = Sample-size adjusted Bayesian information criterion

+ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

^aSmaller SABIC means a better fitting model

^bFixed to zero

6.8 Discussion

In this contribution, we have argued that especially in the field of school improvement research, composition effects should be taken into consideration for the analysis of nested data. And, thus, in multilevel analysis of nested data in school research, it is necessary that the double character of school levels or classroom levels be disentangled as a result of both the global property of a group level – a separate area of responsibility or shared context – and the collective group composition. Furthermore, non-independence and shared higher-level context in nested data do not necessarily result in similar and converging lower level reports – namely, in shared properties – but can also result in a specific configural group property. Therefore, we discussed advances in research on small groups and organizations to present a differentiated model of the double character of group levels in the school environment. We then discussed different types of diversity (separation, variety, and disparity) to describe the composition of a group (in this case, the teaching staff). Methodically, this leads to the necessity of multilevel analyses to include, apart from group means, statistical diversity measures as predictors, such as standard deviation. We then argued that these composition effects could be translated into positioning effects for the individuals of a group because each individual takes a specific position in the composition of a group. The specific individual position can only be described while accounting for the others in the group and in relation to those others. This leads to the methodological proposition of the GAPIM, which provides additional effect terms to conventional multilevel analyses. The others in the group are accounted for with their average values and their similarity among each other as predictors. Further, the relation to those others is accounted for with the similarity of the actor to the others as a predictor. Therefore, the GAPIM allows for the calculation of the effects of the position of individuals within a group regarding an independent variable on an individual dependent variable. We demonstrated the methodological implementation of the GAPIM exemplarily by analysing individual and collective teacher self-efficacy effects on teachers' individual job satisfaction.

The application of the GAPIM has clear advantages over classical multilevel analyses. To begin with, the necessity of multilevel models is usually determined by the presence of a high ICC. The ICC estimates what part of the total variability of a dependent variable is explained by differences between groups and is thus a measurement of the converging influence that a group has on its members. Therefore, with a lower ICC, there would be no assumed nested structure of the data set, and therefore, no further multilevel analysis would be carried out. In our example, a lower ICC was reported regarding job satisfaction, after which further consideration of teaching staff or the group levels would have been obsolete. Including the GAPIM, however, revealed positioning effects that could not be uncovered without considering the nested structure of the data.

The inclusion of the standard deviation as a group composition measure in a multilevel analysis showed no effects of ITE or CTE. In this case, separation of self-efficacy within a group seems to have no effect on the individual level of job

satisfaction. In other words, a teacher's individual job satisfaction does not seem to depend on whether he or she is in a homogeneous or in a highly split teaching staff regarding individual and collective teacher self-efficacy. From a theoretical point of view, it would not have been sensible to conceptualize the diversity of ICE and CTE as variety or disparity. As for other variables in multilevel analyses, Blau's index for variety, or the proportional relation between group members and resources for disparity, could have been included in the same manner as the standard deviation has been. Therefore, this method is promising for formulating questions on different diversity types and providing additional information about composition effects.

Subsequently, the results of the GAPIM showed that position effects of ITE and CTE, indeed, had effects on teachers' individual job satisfaction. In the GAPIM, group composition was translated into position effects by using similarity measures. Similarity measures describe how strongly the actor corresponds with the others in the group regarding the independent variable, as the term I , or how much the rest of the group resembles itself regarding the independent variable, as the term I' .

Regarding ITE, we found that a teacher's job satisfaction was higher, the higher his or her ITE was (main effect of X). However, there is a tendency that job satisfaction was lower, the more the other teachers on staff related to each other regarding their individual self-efficacy (similarity effect of I'), i.e. the homogeneity of the other teachers on staff lowered the measure of influence of individual self-efficacy on job satisfaction (in tendency). *Nota bene*: This effect remained independent, regardless of whether or not the other teachers on staff reported homogeneously high or homogeneously low ITE; it also remained independent, regardless of whether the actor, i.e. a separate teacher, was a part of this homogeneity or not. Since there was no similarity effect I to be found, we have come to know that the similarity of the actor to the other teachers on staff was not important for individual job satisfaction. For individual job satisfaction to occur, it is preferable for a teacher to work together with other teachers who are diverse in their ITE. This becomes transparent, if you consider that, if there is too high homogeneity regarding the individual estimation of ITE, this can limit the possibilities to enter into an exchange with other teachers concerning individual self-efficacy. Individual job satisfaction may decrease, if the rest of a group perceives and acts monolithically.

Regarding CTE, we found that a teacher's individual job satisfaction was higher, the higher collective self-efficacy was as reported by the teacher (main effect of X). In addition, job satisfaction was higher, the more similar the teacher's estimation regarding collective self-efficacy was to the estimation by the rest of the group (similarity effect of I). *Nota bene*: This effect remained independent, regardless of whether or not a teacher's estimated CTE was similarly high or low to his or her colleagues' estimates. Furthermore, the results showed that it was not the average value of the estimations of CTE by the other teachers on staff that had an influence on individual job satisfaction. Therefore, the fact alone that a teacher exhibits a similar estimation as his or her fellow teachers on staff, increases his or her job satisfaction. This can be interpreted as an integration effect. Regardless of how high the estimations are that refer to the shared estimation of CTE, the integration of a shared estimation affects job satisfaction in a positive manner. In contrast, teachers,

who are isolated because of their CTE estimations, show rather low job satisfaction.

Both examples offer arguments supporting the fact that it is not only one's individual and collective teacher efficacy that is of importance for job satisfaction, but also the similarity that prevails within a teaching staff. Yet, the examples imply as well that these similarity effects exhibit complex dynamics. In the case of individual self-efficacy, the similarity of the other teachers on staff decreases a teacher's job satisfaction. This may be explained from a resource-oriented perspective on diversity. Working in a teaching staff, where the other teachers express diverse levels of individual self-efficacy, makes it apparent that individual self-efficacy is alterable and can be affected by different teaching experiences. This could motivate the separate teacher to question work routines and habits and to improve teaching and professionalisation and, thus, lead to higher job satisfaction. In contrast, when the other teachers express a homogeneous level of individual self-efficacy, a teacher could underestimate the possibility of changing work routines and habits and accept his or her individual self-efficacy level as unalterable. Therefore, diversity in individual self-efficacy would be a resource because it serves as a cue to alterable and diverse experiences. In the case of separately perceived collective self-efficacy, the similarity of a teacher to the rest of the teaching staff increases a teacher's job satisfaction. This may be explained from an interference-oriented perspective on diversity. Collective teacher efficacy is meant to be a shared phenomenon and, thus, should be perceived on a similar level by the teachers involved. Therefore, deviations of a separate teacher's perception from the other teachers' perceptions indicate interferences in the group process. Disagreement on a shared foundation can lead to lower job satisfaction.

Therefore, although composition effects on the teaching staff level could not be found, including the GAPIM, research revealed that the composition of a group has an effect on individual job satisfaction through the position of the individual and the individual's similarity relations to the rest of the group. Introducing the GAPIM into school improvement research, then, can provide additional information. Self-evidently, this fact also applies to other unit levels, such as the classroom. Using this method, loneliness and popularity (Gommans et al., 2017; Gommans, Lodder, & Cillessen, 2016) and academic self-concept (Zurbriggen, Gommans, & Venetz, 2016) have been analysed at the classroom level.

6.9 Limitations and Further Research

Despite the theoretically deduced necessity to take composition effects into account, and despite the empirical results that showed that differences between individuals can be explained in a better way by considering additional information on an individual and group level, there are certain difficulties to be expected regarding the implementation of the GAPIM in the field of school improvement research. In field research, we are interested in independent variables that likely have a skewed

distribution. Thereby, it is to be assumed that the multi-collinearity of the different GAPIM terms presents a problem, and this limits the applicability of similarity effects for the analysis. In this contribution, we managed to avoid collinearity by transforming the continuing variables into categorical variables. In addition, the analyses realized in this contribution are limited to cross-sectional data. It would be interesting, for example, to analyse to what extent composition and similarity have an effect on the changes of separate features, e.g. job satisfaction. Further studies need to be conducted in order to examine to what extent dimensions regarding school efficiency and school development are sensitive to composition and similarity effects. Additionally, complementary analyses, such as social network analyses, could increase the benefits of the presented analyses. These analyses are able to make the collective structures and dynamics visible, for example a collective's density or reciprocal relations, and to develop information for the GAPIM regarding the individuals within the collective, for example a person's in- and out-centrality.

In school improvement research, it is widely acknowledged that the school environment has a nested data structure and that diversity within units – in particular within a teaching staff – is of interest. However, this acknowledgment usually does not lead to a differentiated description of how units and groups are composed, what effects such compositions can have, and how such composition effects can be accounted for in statistical methods. In this article, we presented theoretical considerations on the double character of group levels and on the conceptualization of group composition and diversity. In this context, we proposed the methodological advancement of the GAPIM to address this important lack in school improvement research. The example application of the GAPIM to composition and positional effects of individual and collective teacher self-efficacy on job satisfaction showed how the GAPIM can be used in school improvement research and what additional information can be expected.

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York, NY: Free Press.
- Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy beliefs as determinants of teachers' job satisfaction. *Journal of Educational Psychology, 95*, 821–832. <https://doi.org/10.1037/0022-0663.95.4.821>
- Fend, H. (2005). *Neue Theorie der Schule*. Wiesbaden, Germany: Springer.
- Fend, H. (2008). *Schule gestalten: Systemsteuerung, Schulentwicklung und Unterrichtsqualität* (1. Aufl.). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften/GWV Fachverlage, Wiesbaden.
- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*, 479–507.

- Gommans, R., Lodder, G., & Cillessen, A. H. N. (2016). *Effects of classroom likeability composition on adolescent loneliness: A brief introduction to the Group Actor-Partner Interdependence Model (GAPIM)(SRA)*. Utrecht, The Netherlands: Utrecht University Repository.
- Gommans, R., Müller, C. M., Stevens, G. W. J. M., Cillessen, A. H. N., Bogt, T., & Tom, F. M. (2017). Individual popularity, peer group popularity composition and adolescents' alcohol consumption. *Journal of Youth and Adolescence*, *46*, 1716–1726. <https://doi.org/10.1007/s10964-016-0611-2>
- Halbheer, U., Kunz, A., & Maag Merki, K. (2005). *Pädagogische Entwicklungsbilanzen an Zürcher Mittelschulen. Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Lehrpersonen: Schul- und Unterrichtserfahrungen. Skalen- und Iemdokumentation*. Zürich, Switzerland, Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, *9*, 157–191. <https://doi.org/10.1080/0924345980090203>
- Hargreaves, A., & Shirley, D. (2009). *The fourth way: The inspiring future for educational change*. Thousand Oaks, CA: Corwin, a Sage Company.
- Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *The Academy of Management Review*, *32*(4), 1199–1228. <https://doi.org/10.2307/20159363>
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Administrative Science Quarterly*, *44*, 741–763.
- Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits – self-esteem, generalized self-efficacy, locus of control, and emotional stability – with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*, 80–92. <https://doi.org/10.1037/0021-9010.86.1.80>
- Kenny, D. A., & Garcia, R. L. (2012). Using the actor–partner interdependence model to study the effects of group composition. *Small Group Research*, *43*, 468–496.
- Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*, 126–137. <https://doi.org/10.1037/0022-3514.83.1.126>
- Klassen, R. M., Usher, E. L., & Bong, M. (2010). Teachers' collective efficacy, job satisfaction, and job stress in cross-cultural context. *Journal of Experimental Education*, *78*, 464–486. <https://doi.org/10.1080/00220970903292975>
- Kozlowski, S. W. J. (2012). Groups and teams in organizations: Studying the multilevel dynamics of emergence. *Methods for Studying Small Groups: A Behind-the-Scenes Guide*, 260–283.
- Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.
- Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *The Academy of Management Review*, *23*, 325–340.
- LeBreton, J. M., & Senter, J. L. (2007). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*, 815–852. <https://doi.org/10.1177/1094428106296642>
- Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, *33*, 757–798. <https://doi.org/10.3102/00028312033004757>
- Luyten, J. W., & Sammons, P. (2010). Multilevel modelling. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research* (pp. 246–276). New York, NY: Routledge. Retrieved from [https://research.utwente.nl/en/publications/multilevel-modelling\(aff00483-96af-4c5a-95a2-831655655d68\).html](https://research.utwente.nl/en/publications/multilevel-modelling(aff00483-96af-4c5a-95a2-831655655d68).html)

- Maag Merki, K. (Ed.). (2012). *Zentralabitur: Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden, Germany: Springer.
- Maag Merki, K. (2016). Die Einführung zentraler Abiturprüfungen in Bremen und Hessen. In J. Kramer, M. Neumann, & U. Trautwein (Eds.), *Abitur und Matura im Wandel: historische Entwicklungslinien, aktuelle Reformen und ihre Effekte* (pp. 129–159). Wiesbaden, Germany: Springer.
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, 20, 319–350. <https://doi.org/10.1007/s10648-008-9075-6>
- Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997–2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, 34, 410–476. <https://doi.org/10.1177/0149206308316061>
- Merki, K. M., & Oerke, B. (2012). Methodische Grundlagen der Studie. In K. Maag Merki (Ed.), *Zentralabitur: Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland* (pp. 45–61). Wiesbaden, Germany: Springer.
- Mitchell, C., & Sackney, L. (2000). *Profound improvement: Building capacity for a learning community*. Lisse, The Netherlands: Swets & Zeitlinger B.V.
- Moolenaar, N. M., Slegers, P. J. C., & Daly, A. J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education*, 28, 251–262.
- Schmitz, G. S., & Schwarzer, R. (2002). Individuelle und kollektive Selbstwirksamkeitserwartung von Lehrern. *Zeitschrift Für Pädagogik Beiheft*, 44, 192–214.
- Schudel, K. (2012). *Ein Emergenz-basierter Ansatz für die Diversitätsforschung: Soziale Kategorisierung als Reflexion von Verhaltensdiversität*. Unpublished manuscript. https://www.researchgate.net/publication/330116922_Ein_Emergenz-basierter_Ansatz_fur_die_Diversitätsforschung_-_Soziale_Kategorisierung_als_Reflexion_von_Verhaltensdiversitat. <https://doi.org/10.13140/RG.2.2.11603.63529>
- Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin, Germany: <http://www.psyc.de/skalen-doku.pdf>. Accessed 21 Jan 2019.
- Schwarzer, R., & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift Für Pädagogik Beiheft*, 44, 28–53.
- Schwarzer, R., & Schmitz, G. S. (1999). Skala zur Lehrer-Selbstwirksamkeitserwartung. *Skalen Zur Erfassung Von Lehrer- und Schülermerkmalen*, 60–61.
- Schwarzer, R., Schmitz, G. S., & Daytner, G. T. (1999). *The teacher self-efficacy scale*. http://userpage.fu-berlin.de/~health/teacher_se.htm. Accessed 7 Jan 2019.
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, 99, 611–625. <https://doi.org/10.1037/0022-0663.99.3.611>
- Stoll, L. (2009). Capacity building for school improvement or creating capacity for learning? A changing landscape. *Journal of Educational Change*, 10, 115–127. <https://doi.org/10.1007/s10833-009-9104-3>
- Van Knippenberg, D., de Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, 89, 1008–1022. <https://doi.org/10.1037/0021-9010.89.6.1008>
- Van Knippenberg, D., & Schippers, M. C. (2006). Work group diversity. *Annual Review of Psychology*, 58, 515–541. <https://doi.org/10.1146/annurev.psych.58.110405.085546>
- Zurbriggen, C., Gommans, R., & Venetz, M. (2016). *The big-fish-little-pond effect on academic self-concept: A comparison of GAPIM and a latent-manifest contextual model (SRA)*. Utrecht, The Netherlands: Utrecht University Repository.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

