A New Instrument for Assessing Cohesion in Primary and Grammar School Classes—Factorial Structure and Measurement Invariance of the GruKo4

Sina Schürer1, Stefanie van Ophuysen1, and Lars Behrmann1

Abstract
Class climate has been focused in the context of school research for decades. One central facet of class climate is cohesion. Whereas there are well-elaborated instruments to assess cohesion in different contexts (e.g., sport and therapy), such an instrument is missing for the school context. The aim of the current article is to present an instrument to measure different facets of group cohesion in classes of primary and grammar schools. The factorial structure of the instrument is analyzed using data from Grades 2 to 7 (N ≈ 3052 from 146 classes) by means of confirmatory factor analysis. Furthermore, tests of measurement invariance across the different levels of education (primary vs. grammar school) are undertaken. Results support configural as well as partial metric and scalar invariance between students of primary and grammar schools. Further, cohesion in primary schools is significantly higher than in grammar schools. Limitations of the study and implications for future research in the context of school are discussed.

Keywords
cohesion, school climate, measurement invariance, scale development/testing, measurement

Cohesion is one aspect of the school climate. However, in contrast to other contexts, such as sports, the construct of cohesion has so far received little attention in school contexts. There is no elaborate instrument that maps the various facets of cohesion on the basis of a well-founded theory. The main purpose of the current study is to present a comprehensive instrument for assessing cohesion in school classes from Grades 2 to 7 and to investigate measurement equivalence across different school levels in Germany.

1WWU Münster, Münster, Germany

Corresponding Author:
Sina Schürer, AG Forschungsmethoden, Westfälische Wilhelms-Universität Münster, Georgskommende 33, Münster 48149, Germany.
Email: schuerer@uni-muenster.de
Cohesion

Four Facets of Group Cohesion

Festinger (1950, p. 274) defines cohesion as “the resultant of all forces acting on members of groups to remain in the group.” Cohesiveness is an important group characteristic; among other things, cohesive groups help their members to fulfill the basic human need for belonging (Correll & Park, 2005; Johnson et al., 2006), which is defined as “a pervasive drive to form and maintain at least a minimum quantity of lasting, positive, and significant interpersonal relationships” (Baumeister & Leary, 1995, p. 497). As meta-analyses have shown, there are several negative emotional consequences, including loneliness, depression (Blackhart, Nelson, Knowles, & Baumeister, 2009), and reduced self-esteem (Gerber & Wheeler, 2009), if this need is not met.

In his definition of cohesion, Festinger stressed that cohesion depends “on the attractiveness or unattractiveness of either the prestige of the group, members in the group, or the activities in which the group engages” (Festinger, 1950, p. 274). Consequently, an essential characteristic of a cohesive group is that its members feel attracted to it (= attraction to the group, ATG). Another key feature of a cohesive group is that its members feel strongly connected to each other; a cohesive group itself is highly integrated (= group integration, GI). Early on, Van Bergen and Koekbakker (1959, p. 366) have therefore stated that a cohesive group is characterized by “a closeness among members, a similarity in the perception of events, and [...] a bonding together in response to the outside world.” Taken together, two characteristics of cohesion emerge: (1) from the perspective of the individual group member, cohesion implies that each member feels strongly ATG. (2) From the perspective of the group itself, cohesion means that group members are perceived as being strongly connected (GI).

A more recent definition describes cohesion as “the tendency for a group to stick together and remain united in the pursuit of its instrumental objectives and/or for the satisfaction of member affective needs” (Carron, Brawley, & Widmeyer, 1998, p. 213). This definition additionally draws attention to the instrumental functions of a group. This might be the fulfillment of a common task or reaching a common goal and/or the fulfillment of the affective need for belonging (social participation). Thereby, they distinguish between a task-related and a social function.

Finally, Carron, Widmeyer, and Brawley (1985) differentiate four facets of group cohesion resulting from the combination of the two perspectives (ATG vs. GI) with the two group functions (social vs. task) (Carron et al., 1985; Carron & Brawley, 2000):

1. ATG-social: a desire to belong to the group because of the people.
2. ATG-task: a desire to belong to the group because of the tasks/goals.
3. GI-social: perception of tight social relations between people in the group.
4. GI-task: perception of a good working atmosphere.

The differentiation of these facets has already proven to be useful, especially in nonschool contexts. In a validation study of the group environment questionnaire (GEQ), a tool for assessing the four facets of cohesion in the area of sports, the mean differences between athletes from team sports versus individual sports differed depending on the facet of cohesion considered. While GI-task and ATG-task each differed significantly in the expected direction, no significant differences were found for the two social facets. However, the two social facets were predictive for the duration of team membership for individual athletes (Brawley, Carron, & Widmeyer, 1987). In a correlational study with a German adaptation of the GEQ, GI-task and ATG-task correlated more strongly than the two social facets with three satisfaction items concerning the coach, communication, and the team as a whole (Ohlert, 2012). The same instrument was used in a study with participants of rehabilitation sports groups. Differential correlations were also uncovered here:
The fulfillment of the need for relationships correlated most strongly with ATG-social, while the attitude toward sports activity correlated most strongly with ATG-task (Kleinknecht, Kleinert, & Ohlert, 2014). The four-facet model was also used to design team-building interventions. Carron and Spink (1993), for example, developed a successful intervention to strengthen the cohesion of fitness groups by systematically designing measures to strengthen the four facets of the model.

Cohesion in the School Context

In the context of school research, a comprehensive concept of cohesion is missing so far. Rather, cohesion is regarded as one aspect of the broader construct of class climate. Often, it is simply described as the quality of students’ connectedness (Eder, 1996). However, research on predictors and consequences of class cohesion should be more beneficial, if based on a more well-developed understanding of the concept. The cohesion model developed by Carron and Brawley (2000) can be applied to any group with a common identity, common goals, and common objectives, that is, groups which exhibit structured patterns of interaction and communication, whose members are interdependent both socially and while completing tasks, and who feel interpersonal attraction toward each other and self-identify as a group. School classes fulfill all these criteria. Students in classes cooperatively work toward goals and tasks and share experiences, thereby qualifying as interdependent. A class is also distinguishable from other groups/classes by external conditions (e.g., name of the class or the classroom), so pupils within a class can perceive themselves as a part of a distinct group.

Applying the model to the school context results in four facets of cohesion (see also Table 1). From the perspective of an individual child, the “attractiveness” of the class can be either social or task-based:

- **ATG-social**: This facet relates to social interactions between classmates including extracurricular activities or playing together in the breaks. How socially attractive the class is depends on the affective quality of the peer relationships which develop.
- **ATG-task**: In relation to task-based or cognitive-instrumental relations (Zander, Kreutzmann, & Hannover, 2017), a group will be very attractive if the children enjoy learning and working together and enjoy the lessons themselves. Thus, a high task-related attractiveness results when the child positively evaluates the activities in class and wants to participate in them.

Just like attraction, connectedness or “GI” can also be judged either focusing on social or task-related functions of the group:

- **GI-social**: The social aspect relates to the perception of commonalities among class members and the level of perceived connectedness individual children experience in social, extracurricular situations. It is about the presence of close social relationships between the children in the class and the degree to which they perceive themselves as similar and connected to their peers and how well the group holds together.
- **GI-task**: Task-based cohesion relates to how well the group functions as a learning and working community. A high level of GI-task reveals good networking when it comes to activities related to learning and working together, for example, helping and supporting each other while completing tasks.

The distinction between the two group functions, task and social, is particularly meaningful for the study of school classes. On the one hand, school classes allow for cognitive-instrumental peer relations by joint learning and achieving common learning goals (task). On the other hand, they offer space for affective relations with peers and thereby fulfilling the students’ need to belong and social contact (social) (Zander et al., 2017). Some teachers might invest more time in developing
a strong community with a corporate feeling by undertaking common class trips or team building activities. Other teachers, in contrast, focus more on collective performance and a cooperative working atmosphere. Therefore, the degree of task-based or social cohesion may vary independently between classes (e.g., depending on teachers’ actions), and moreover, the consequences for students’ achievement and well-being may be related in different ways to these cohesion facets.

Also, as Carron and Brawley (2000) stressed, group cohesion can be assessed from different perspectives: (1) a cohesive class is one whose students consider the class attractive and (2) in a cohesive class, students experience strong networking. This distinction seems plausible: It could be the case that the pupils want to belong to a class, although the children in that class are not well connected (high ATG but low GI), and it is also feasible that a class is not attractive for some children, although the connection of the other group members is high (low ATG but high GI). These explanations demonstrate that applying this cohesion model to the school context is both possible and reasonable.

### Cohesion in Primary and Grammar School Classes

Class cohesion is an equally interesting topic in primary (Grades 1–4 in most German federal states) and secondary education. Students of these two school levels not only differ in age but also experience typical differences in the way lessons are organized and designed. It is precisely these different experiences that should be accompanied by systematic differences in class cohesion, which in turn should vary according to the cohesion facets under consideration. In primary school, for example, cooperation with partners or in small groups is more common than in secondary school. In contrast, individual performance is more strongly emphasized in secondary school—at least in grammar schools (Gymnasium)—which is the type of school with the highest performance requirements. Therefore, we expect a higher task-related GI in primary school than in grammar school. However, we do not expect differences in the social dimension of GI. We also expect differences in the task-related, but not the social dimension of the ATG. As already well-documented, liking of school and learning motivation decline with the age of the school children (e.g., Eccles et al., 1993; Hagenauer & Hascher, 2011; Helmke, 1997). Therefore, working and learning together should be less attractive for grammar school students than for younger students at primary school. An adequate instrument for measuring school cohesion should be able to reflect these school type-related differences.

### Measurement of Group Cohesion

Carron and colleagues developed instruments for the assessment of the four facets of cohesion in sports groups for adults (GEQ, Brawley et al., 1987) and adolescents (Youth sport environment
questionnaire (YSEQ) Eys, Loughead, Bray, & Carron, 2009). Scales with multiple items represent each of the four facets. Aggregating the individual scale values at the group level produces four cohesion indices with moderate correlations (e.g., Carron et al., 1985, \( r = .32 - r = .54 \)).

Empirical research with these instruments largely confirmed this concept of cohesion not only for teams in competitive, leisure, and health sports (see Eys et al., 2009; Kleinknecht et al., 2014; Martin, Carron, Eys, & Loughead, 2012; Ohlert, 2012) but also for groups of musicians, working groups, or therapy groups (e.g., Carless & Paola, 2000; Dyce & Cornell, 1996; Estabrooks & Carron, 2000; Li & Harmer, 1996; Whitton & Fletcher, 2014).

The differentiation of the two functions (social vs. task) was confirmed for different contexts (sports teams, working groups, and therapy groups) and different age groups—especially for adolescents (13–18 years) and children (9–12 years) (Eys et al., 2009; Martin et al., 2012). Concerning the empirical differentiation of the two perspectives (ATG vs. GI), empirical evidence is ambiguous. While some studies could not replicate a clear differentiation (Carron et al., 1985; Li & Harmer, 1996; Whitton & Fletcher, 2014), other studies could confirm the four facets (Carless & Paola, 2000; Dyce & Cornell, 1996; Schutz, Eom, Smoll, & Smith, 1994). Particularly for sports groups of children and adolescents, the factorial differentiation between ATG and GI could not be empirically verified (Eys et al., 2009; Martin et al., 2012).

### Previous Instruments for Measurement of Group Cohesion in the School Context

As already mentioned, cohesion has been used as part of the class climate in school-related research. As such, cohesion is represented as a subscale in various instruments of classroom climate assessment. For example, both the widely used learning environment inventory (LEI, Fraser, Anderson, & Walberg, 1982) and its simplified version, the My Classroom Inventory (MCI), contain a subscale called cohesion that assesses the degree that “students, know, help, and are friendly toward each other” (Fraser et al., 1982, p. 5). The seven items of the cohesion scale from the LEI all represent the GI-social facet. The cohesion subscale of the MCI consists of six items, four of which can also be assigned to the GI-social facet. However, two items capture a different construct, namely, individually perceived social participation (e.g., “In my class everybody is my friend,” Fraser et al., 1982, p. 51). Also in a revised, shortened version of the MCI (MCI-SF, Fraser & Fisher, 1983), the GI-social facet of cohesion is mixed up with items of perceived social participation. A further development of the LEI is the Science Laboratory Environment Inventory (Fraser, McRobbie, & Giddings, 1993), which also contains a subscale called cohesion. These items ask for the scope of mutual knowledge, help, and support in scientific courses (e.g., “Members of this laboratory class help one another,” Fraser et al., 1993, p. 18) and therefore focus on task-related rather than social cohesion. The cohesion subscale in an instrument by Oswald (1989, p. 82) is even more diverse and covers aspects of mutual assistance and cooperation (task) as well as of belonging, trust and community (social) from the individual perspective (attraction) as well as from the group perspective (GI).

Taken together, existing instruments are not suitable for a differentiated assessment of class cohesion. Concerning the school climate in general, Anderson (1982) posed the question “Are we all hunting the same beast?” (p. 376). A look at the existing scales also raises this question for cohesion and illustrates the need for an instrument that is based on an elaborate, theoretically sound construct.

### Purpose of this Study

The first purpose of this study is to examine the factorial structure of a new instrument for the assessment of group cohesion, the GruKo4.1. Moreover, we test for measurement invariance of the
instrument with regard to the school type (primary school vs. grammar school). If measurement invariance holds, the differences of the specific cohesion facets between the two school levels are further examined. We assume that in primary schools, the task-related facets of cohesion are more pronounced than in grammar schools, while differences should be small concerning the social facets.

**Method**

**Participants**

The analyses are based on the data of students from Grades 2 to 4 from 20 North Rhine-Westphalian (NRW) primary schools ($N = 1684$, 78 classes) as well as Grade 7 students from 26 grammar schools ($N = 1368$, 68 classes). Students in grade 7 belong to the highest age group for which the instrument is suitable and therefore the biggest differences to primary school children can be expected. Data collection in primary schools took place as part of the project SoPaKo (OP 158/4–1, Social Participation through Cohesion, $N = 11$ schools) as well as in a preliminary study on this project ($N = 9$ schools). The participating grammar schools were schools from the research project GanzIn (a project on all-day schools in NRW). Table 2 shows more detailed characteristics of the sample description.

In both samples, the migration background is operationalized via the respective country of birth of the children. Overall, the proportion of students who were not born in Germany is very low. Especially at the grammar schools, only 4.5% of children are not born in Germany (Table 3).

**Table 2.** Description of the Two Samples.

| School form | Grade | $N$ | Of which boys | $M$ age ($SD$) |
|-------------|-------|-----|---------------|----------------|
| Primary     | 2     | 633 (37.9%) | 310 (49.0%) | 7.13 (.530) |
|             | 3     | 874 (52.9%) | 442 (50.6%) | 8.23 (.555) |
|             | 4     | 154 (9.1%)  | 70 (45.5%)  | 9.55 (.595) |
|             | Total | 1661 | 822 (49.5%)  | 7.94 (.913) |
| Grammar     | 7     | 1365 | 627 (47.9%)  | 12.54 (.529) |

**Table 3.** Description of Migration Background of the Two Samples.

| School   | Born in Germany | Not born in Germany | Children total |
|----------|-----------------|---------------------|---------------|
| Primary  | 1468 (88.9%)    | 183 (11.1%)         | 1651          |
| Grammar  | 1302 (95.5%)    | 62 (4.5%)           | 1364          |

**Instrument**

The GruKo4 consists of four subscales, which describe the four facets of cohesion from the model developed by Carron et al. (1985). It is based on instruments from the sports sector (e.g., GEQ, Brawley et al., 1987; YSEQ, Eys et al., 2009), but the choice of words has been adapted to the school context. With regard to the age of the pupils, items were formulated in simple language. The four scales with example items are listed below:

1. ATG-social: “I like the children in my class.”
2. ATG-task: “I like going to class.”
3. GI-social: “In my class, we all stick well together.”
4. GI-task: “My class is a really good learning community.”
The ATG subscales each consist of five items, the GI subscales of three items each. Each item is answered on a four-level rating scale (“exactly true” to “not at all”).

**Procedures**

The data collection in primary schools was carried out as a part of the preliminary study of the SoPaKo project at the beginning of 2017, and further, data were collected in September 2017 as a part of the first measurement phase of the SoPaKo project. The GruKo4 scales were a part of a larger psychometric questionnaire in both cases. Each child completed its own questionnaire. The third and fourth grade children completed theirs in their respective class environment, while the second graders were broken up into smaller groups. Trained test leaders read each question out loud, while the children listened quietly and ticked their corresponding answer. Data were collected from the secondary school students as part of the second survey phase of the GanzIn project. A part of this survey was completed online (\(N = 1164\)) and part using paper-and-pencil (\(N = 204\)). A multigroup comparison showed no significant factorial differences between the two groups (\(\Delta \text{CFI} = .007, \Delta \text{RMSEA} = .015\)).

**Data Analyses**

The statistical analyses were carried out using software Mplus 7 (Muthén & Muthén, 1998–2018) and with SPSS 25 (Statistical Package for Social Science, IBM®). Missing data were handled by Mplus using full information maximum likelihood estimation. The command “type = complex” basically considered the nested structure of the data and used the class as a cluster variable.

First of all, a confirmatory factor analysis was carried out for the overall sample to check whether the theoretically assumed four-factorial structure can be mapped. The four-factorial model was compared with a one-factor solution (one total factor of cohesion). To evaluate model fit, multiple fit-statistics were used: \(\chi^2\), comparative fit index (CFI), and root mean square error of approximation (RMSEA). CFI statistics \(\geq .95\) and RMSEA statistics \(\leq .05\) are denoted as “good” fit (Geiser, 2011). Hu and Bentler (1999) suggest values close to .06 for RMESA statistics to conclude that there is a relatively good fit. Hair, Babin, Anderson, and Black (2014) associate CFI values \(\geq .90\) with a good model fit. According to Chen (2007), CFI and RMSEA are sensitive to lack of invariance. The \(\chi^2\) difference test compares the models. Since the values are not normally distributed, the Satorra–Bentler correction procedure was applied.

Following the factor analysis, descriptive parameters (mean values, SDs, skewness, and kurtosis) for each item are reported for both samples. Also, the internal consistency of the four scales was examined by calculating Cronbach’s alpha.

To check the measurement invariance, a stepwise procedure was chosen (see Christ & Schlüter, 2012). It started with the least restrictive model, followed by comparisons with the more restricted models. The weakest form of measurement invariance is the *configurational invariance*. This is guaranteed as soon as the factorial structure is the same in both groups, which means that the same items load on the same factors, and goodness of fit indices are proving good for both groups. Weak factorial invariance or *metric invariance* is given when factor loadings in both groups are the same. Besides, strong factorial or *scalar invariance* can be assumed in the case of equal intercepts, which is necessary for comparing the mean values in the two groups. To test the measurement invariance, \(\chi^2\) difference tests are used to gradually compare models with increasing constraints. If no significant deterioration of the model fit is detectable, the more restrictive form of invariance is accepted. For larger samples, however, even very small model deviations lead to significant results in the test. Therefore, when sample size is large, Chen’s “rule of thumb” can be applied. According to this rule, differences between model fits can be considered irrelevant, if \(\Delta \text{CFI} < .01\) and
ΔRMSEA is <.015 (e.g., Chen, 2007; Cheung & Rensvold, 2002). In cases where both models fit equally well statistically, the more restrictive model can be accepted. Finally, we examined the differences in the latent mean values of the four cohesion facets between the two samples (primary school vs. grammar school).

**Results**

**Confirmatory Factor Analysis**

Results of the first CFAs confirm the hypothesized 4-factor model of cohesion. Compared to a 1-factor model, the fit-statistics of the 4-factor model fit the data better (see Table 4). The values of CFI and RMSEA indicate an inappropriate model fit of the 1-factor model. The $\chi^2$ difference test reveals a significantly better fit of the 4-factor model (Satorra–Bentler scaled $\Delta \chi^2 = 1909.3035$, df = 6, and $p < .001$). Therefore, the 4-factor solution is assumed now and in the following.

Figure 1 illustrates the 4-factor solution, including correlations between the subscales and factor loadings. As expected, the four subscales of cohesion are not completely independent of each other.

### Table 4. Results of the Confirmatory Factor Analysis of the GruKo4.

|                     | CFI  | RMSEA | $\chi^2$ |
|---------------------|------|-------|----------|
| 1-factor model      | .742 | .112  | 4071.488 |
| 4-factor model      | .967 | .042  | 612.810  |

*Note. CFI = comparative fit index; RMSEA = root mean square error of approximation.*

![Figure 1. GruKo4, the 4-factor solution of confirmatory factor analysis with factor loadings and correlations between the four facets.](image)
Descriptive Results and Reliability

Descriptive statistics are summarized in Table 5. The value for skewness indicates that the distribution is left-aligned. The internal consistency proved to be good for all four subscales in both samples ($\alpha = .752$–$\alpha = .927$).

Table 5. Descriptive Item Statistics and Reliability for Each Sample and Subscale.

| Scale and Item                             | Primary M (SD) | Skew (SE) | Kurt (SE) | Grammar M (SD) | Skew (SE) | Kurt (SE) |
|-------------------------------------------|----------------|-----------|-----------|----------------|-----------|-----------|
| GI-social                                 |                |           |           |                |           |           |
| In my class, we all stick well together   | 3.46 (.66)     | -1.58 (.06) | 1.94 (.12) | 2.86 (.83)     | -.41 (.07) | .65 (.13) |
| In my class, we all pay a lot of attention to each other | 3.40 (.69)     | -1.37 (.06) | 1.22 (.12) | 2.78 (.79)     | -.25 (.07) | .70 (.13) |
| In my class, we all fit together really well | 3.21 (.86)     | -.99 (.06)  | .04 (.12)  | 2.70 (.81)     | -.21 (.07) | .72 (.13) |
|                                            | $\alpha = .789$ |           |           | $\alpha = .872$ |           |           |
| GI-task                                   |                |           |           |                |           |           |
| In my class, we all work well together    | 3.35 (.80)     | -1.28 (.06) | .67 (.12)  | 2.81 (.70)     | -.39 (.07) | .34 (.13) |
| In my class, we all make sure everyone can do their jobs | 3.33 (.83)     | -1.22 (.06) | .49 (.12)  | 2.46 (.73)     | .01 (.07)  | .64 (.13) |
| My class is a really good study community | 3.35 (.85)     | -1.31 (.06) | .67 (.12)  | 2.51 (.84)     | -.02 (.07) | .82 (.13) |
|                                            | $\alpha = .752$ |           |           | $\alpha = .824$ |           |           |
| ATG-social                                 |                |           |           |                |           |           |
| I like the kids in my class               | 3.55 (.48)     | -1.66 (.06) | 2.73 (.12) | 3.16 (.61)     | -.81 (.07) | .48 (.13) |
| I think the kids in my class are nice      | 3.50 (.55)     | -1.58 (.06) | 2.26 (.12) | 3.15 (.62)     | -.80 (.07) | .51 (.13) |
| After the holidays I look forward to the children of my class | 3.56 (.62)     | -1.92 (.06) | 3.03 (.12) | 3.09 (.78)     | -.74 (.07) | .19 (.13) |
| I like being with the kids in my class     | 3.56 (.55)     | -1.79 (.06) | 2.79 (.12) | 3.10 (.67)     | -.75 (.07) | .17 (.13) |
| I’d be sad if I couldn’t be with the kids in my class anymore | 3.52 (.84)     | -1.85 (.06) | 2.14 (.12) | 3.10 (.78)     | -.74 (.07) | -.21 (.13) |
|                                            | $\alpha = .752$ |           |           | $\alpha = .824$ |           |           |
| ATG-task                                   |                |           |           |                |           |           |
| I like the things we do in class           | 3.35 (.64)     | -1.24 (.06) | 1.17 (.12) | 2.52 (.62)     | -.12 (.07) | .42 (.13) |
| I like to learn new things in class        | 3.62 (.53)     | -2.13 (.06) | 4.19 (.12) | 2.68 (.62)     | -.28 (.07) | .28 (.13) |
| After the holidays I’m looking forward to my lessons | 2.99 (1.26)     | -.69 (.06)  | -.97 (.12) | 2.11 (.91)     | .43 (.07)  | .81 (.13) |
| I like going to class                      | 3.14 (.93)     | -.90 (.06)  | -.23 (.12) | 2.39 (.78)     | .05 (.07)  | .75 (.13) |
| I’d be sad if I couldn’t go to class anymore | 3.15 (1.28)     | -.96 (.06)  | -.61 (.12) | 2.51 (.88)     | -.09 (.07) | .88 (.13) |
|                                            | $\alpha = .766$ |           |           | $\alpha = .927$ |           |           |

Note. Skew = skewness; Kurt = kurtosis; SE = standard error of each item; ATG = attraction to the group; GI = group integration.
**Measurement Invariance Analysis**

In the first step, CFAs were calculated separately for both samples. In both cases, the models revealed a good fit (see Table 6) so that configural measurement invariance (CMI) can be assumed. Our assignment of items to the four (correlated) subscales fits well to the data structure (primary school: CFI = .963, RMSEA = .034; grammar school: CFI = .954, RMSEA = .063).

|                      | CFI   | ΔCFI  | RMSEA | ΔRMSEA | $\chi^2$ (df) | $\Delta\chi^2$ (df) |
|----------------------|-------|-------|-------|---------|----------------|---------------------|
| Primary school       | .963  | .034  | .051  | .003    | 286.773 (98)   | 140.82 (16)         |
| Grammar school       | .954  | .063  | .049  | .001    | 628.501 (98)   |                     |
| Baseline model       | .956  | .048  | .053  | .004    | 894.536 (196)  |                     |
| MMI                  | .948  | -.008 | .051  | .003    | 1035.356 (212) | 140.82 (16)         |
| PMI                  | .950  | -.006 | .049  | .001    | 994.582 (211)  | 100.05 (15)         |
| SMI                  | .939  | -.011 | .053  | .004    | 1193.077 (223) | 198.50 (12)         |
| PSMI                 | .941  | -.009 | .052  | .003    | 1149.491 (222) | 154.91 (11)         |

Note. Free factor loading for ATG-t2 for PMI and SMI; additionally free intercept for ATG-t2 for PSMI. CFI = comparative fit index; RMSEA = root mean square error of approximation; MMI = metric measurement invariance; PMI = partial measurement invariance; SMI = scalar measurement invariance; PSMI = partial scalar measurement invariance; ATG = attraction to the group.

Next, a baseline model was calculated for later comparison with the following more restrictive measurement models. In the baseline model, the model parameters of the groups to be compared are freely estimated. The model fit is good (CFI = .956, RMSEA = .048).

In the next step, metric measurement invariance was tested. For this purpose, factor loadings were restricted to be equal in both groups. Then, model fit indices were compared with those of the CMI model. Although fit values remained satisfactory, a significant decline resulted in comparison to the fit of the CMI model ($\Delta\chi^2 = 140.82$, $\Delta df = 16$, $p <.001$). Statistically significant differences between the two groups have to be assumed in at least one factor loading. Using Chen’s rule of thumb, it is revealed that the CFI decline is less than .01 units ($\Delta\text{CFI} = -.008$) and the RMSEA increases by less than .015 units ($\Delta\text{RMSEA} = +.003$). Thus, from a more practical perspective, the assumption of metric invariance does not necessarily have to be rejected. Nevertheless, a look at the modification indices shows that there is an improvement in the model, if the factor loading of the ATG-t2 item is freely estimated. After the recalculation of the model with free factor loading for item ATG-t2, the change is still significant ($\Delta\chi^2 = 100.05$, $\Delta df = 15$, but $p <.001$), but there are only minimal changes in the fit indices ($\Delta\text{CFI} = -.006$, $\Delta\text{RMSEA} = +.003$) compared to the baseline model.

In a final step, the same test was carried out for scalar measurement invariance. For this purpose, in addition to factor loadings (except ATG-t2), the intercepts of the manifest variables between the two groups were equated. Again, there is a statistically significant decrease of the model fit ($\Delta\chi^2 = 198.50$, $\Delta df = 12$, $p <.001$). This decrease is supported by a negative change in CFI ($\Delta\text{CFI} = -.011$), which is slightly above the acceptable change ($\Delta\text{CFI} <.01$). However, the increase in RMSEA is rather small ($\Delta\text{RMSEA} = +.004$). Considering these results, if the factor loading and intercept of the item ATG-t2 are freely estimated, then the model degradation is still significant ($\Delta\chi^2 = 1149.491$, $\Delta df = 11$, $p <.001$), but the differences in the other two fit values fall within acceptable limits ($\Delta\text{CFI} = -.009$, $\Delta\text{RMSEA} = +.003$).

### Comparison of the Perception of the Cohesion Facets for the Subsamples

The test for measurement invariance shows that partial scalar measurement invariance can be assumed. For the comparison of cohesion across the school levels, item ATG-t2 is excluded. To compare the four
cohesion facets in the different school levels, the latent mean values of the subscales of the secondary level were compared with the values of the primary level (0). For all four scales, there are significant differences between the two groups (see Table 7). Latent mean values are lower in grammar school classes than in primary school classes, suggesting that students in primary school exhibit higher cohesion. In particular, the two task facets show large differences in the extent of cohesion.

| Table 7. Deviation of Grammar School Latent Mean Values. |
|---------------------------------------------------------|
| Grammar school                                        |
| Latent M (SE)                                          |
| GI-s                                                   | −0.82 (0.04) |
| GI-t                                                   | −1.15 (0.05) |
| ATG-s                                                  | −0.68 (0.04) |
| ATG-t                                                  | −1.12 (0.04) |

Note. ATG = attraction to the group; GI = group integration.

Discussion

The main objective of the current study was to present a new instrument for assessing the four facets of cohesion from the model of Carron et al. (1985) in school classes and to investigate measurement equivalence across different school levels. It has been revealed that the four facets could be measured reliably and that they correlate in moderate strength. They thus proved to be not only theoretically but also empirically distinguishable facets of a common underlying construct. Furthermore, the GruKo4 scales revealed to be measurement invariant with regard to the school type. They can be used to analyze the differences between primary school and grammar school. In this respect, students from grammar school showed lower scores on average in all four facets of cohesion. As expected, the differences were greater in the task-related facets than in the social facets of cohesion.

With the GruKo4, the four facets of cohesion from the model of Carron et al. (1985) can be measured at the school context with students from Grades 2 to 7. This opens the way to interesting new research questions. The separation of task-based and social aspects can be particularly informative, for example, if the relationship between the design of teaching and extracurricular activities and cohesion is being examined. Beyond these class-wise predictors, the analysis of individual consequences of the four different facets of cohesion—for example, in terms of achievement, school liking, well-being, social participation or exclusion, and bullying—seems an interesting research field.

In the extracurricular sector, connections between group cohesion and performance have been repeatedly confirmed. Both in the context of working groups and sports groups, it was shown that cohesive groups perform better (Beal, Cohen, Burke, & McLendon, 2003; Carron, Colman, Wheeler, & Stevens, 2002). A similar connection can also be assumed in the school context because in a cohesive class, students should support each other in their learning and working. Although research in the sports context has so far failed to support the theoretically well-founded assumption of differential correlations between the different facets of cohesion and performance (Carron et al., 2002), school seems to be a good candidate to reexamine this assumption. Relevant questions might be: Is task-related cohesion more important than social cohesion? Is the attractiveness of tasks (ATG) more important than task-related cohesion (GI)?

Furthermore, there are first indications of the connection between cohesion and social participation (Schürer, 2019). Social participation goes hand in hand with the fulfillment of the need to belong. If this need is not satisfied, then there are serious negative consequences for the
individual. Thus, cohesion in classrooms might become increasingly important in the current society and in increasingly heterogeneous classrooms, where exclusion of minorities occur far too frequently. The different effects of the task-related and social facets of cohesion on the social participation of individual students in their class have not yet been investigated and appear to be a valuable research question.

Based on this knowledge, intervention studies can be designed in order to improve the overall cohesion or specific facets of cohesion. Within the framework of the SoPaKo project, from which a part of the data presented here originates, a complex intervention was developed which consists of learning-based as well as social activities in dyads or groups of students during regular instruction. These individual measures were each tailored to a specific facet of cohesion (Schürer, 2019). However, since the intervention was implemented as a whole, it has not yet been possible to verify the specific effectiveness of individual measures and is awaiting further studies. The GruKo4 provides an instrument for the evaluation of such interventions which seek to foster the specific facets of cohesion.

Since the instrument turned out to be measurement invariant in relation to the type of school, we checked the differences in the mean values for the four cohesion facets between primary and grammar schools. We found significant differences between the two samples for all four latent variables. As expected, the differences were stronger for the two task-related facets of cohesion. This result provides initial evidence of validity of the instrument and a justification for its use in future studies. However, further studies to validate the instrument are still pending. Correlations of the GruKo4 scales with cohesion scales from classroom climate instruments (e.g., LEI and MCI) could be one obvious approach.

Apart from the need for further validation studies, it is also necessary to expand the sample in order to allow a broader generalization of the findings. One limitation of our study is that all secondary school students attended a Grade 7 class at one of the all-day grammar schools in NRW. The grammar school is the highest track in the German school system. On average, students at this track differ from the overall student population not only in their academic achievement but also in terms of socioeconomic status, migration, and educational background (e.g., Hammer et al., 2016; Rauch, Mang, Härtig, & Haag, 2016; Schiepe-Tiska et al., 2016). Therefore, the results of our study should not be generalized to the other types of secondary schools without further investigation. The focus on Grade 7 students is, in our opinion, a minor problem at least with respect to the question of measurement invariance. Seventh graders are the oldest students for whom the questionnaire should be used since some item formulations do not seem appropriate for older students. Thus, the invariance could be assigned with (age-wise) maximally different subsamples. It therefore seems obvious that measurement invariance should also apply to younger students at grammar schools. Nevertheless, it would be worthwhile to extend the sample to Grades 5 and 6 in further studies not least because the differences in the mean values may depend more on age than on school type.

So far, we can state that GruKo4 is a theoretically sound instrument for the differentiated assessment of cohesion among students from Grades 2 to 7 in primary schools and grammar schools. It offers the possibility for longitudinal analysis of cohesion changes, for example, in the context of primary school transition as one prominent example, and seems useful for the evaluation of pedagogical interventions that shall foster cohesion in general—for example, team building training—or its special facets—for example, cooperative learning or school trips.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Deutsche Forschungsgemeinschaft (Grant: OP 158/4-1).

ORCID iDs
Sina Schürer  https://orcid.org/0000-0003-1256-6899
Stefanie van Ophuysen  https://orcid.org/0000-0002-7829-3363

Note
1. The “GruKo” in GruKo4 stands for the German word Gruppenkohäsion meaning group cohesion. 4 relates to the instrument’s four facets.

References
Anderson, C. S. (1982). The search for school climate: A review of the research. Review of Educational Research, 52(3), 368-420. doi:10.3102/00346543052003368
Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychological Bulletin, 117(3), 497-529.
Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. Journal of Applied Psychology, 88(6), 989-1004. doi: 10.1037/0021-9010.88.6.989
Blackhart, G. C., Nelson, B. C., Knowles, M. L., & Baumeister, R. F. (2009). Rejection elicits emotional reactions but neither causes immediate distress nor lowers self-esteem: A meta-analytic review of 192 studies on social exclusion. Personality and Social Psychology Review, 13(4), 269-309. doi:10.1177/1088868309346065
Brawley, L. R., Carron, A. V., & Widmeyer, W. N. (1987). Assessing the cohesion of teams: Validity of the group environment questionnaire. Journal of Sport Psychology, 9, 275-294.
Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. Small Group Research, 31(1), 71-88.
Carron, A. V., & Brawley, L. R. (2000). Cohesion: Conceptual and measurement issues. Small Group Research, 31(1), 89-106. doi:10.1177/104649640003100105
Carron, A. V., Brawley, L. R., & Widmeyer, W. N. (1998). The measurement of cohesiveness in sport groups. In J. L. Duda (Ed.), Advances in sport and exercise psychology measurement (pp. 213-226). Morgantown, WV: Fitness Information Technology.
Carron, A. V., Colman, M. M., Wheeler, J., & Stevens, D. (2002). Cohesion and performance in sport: A meta-analysis. Journal of Sport and Exercise Psychology, 24, 168-188.
Carron, A. V., & Spink, K. S. (1993). Team building in an exercise setting. The Sport Psychologist, 7(1), 8-18.
Carron, A. V., Widmeyer, W. N., & Brawley, L. R. (1985). The development of an instrument to assess cohesion in sport teams: The group environment questionnaire. Journal of Sport Psychology, 7(3), 244-266.
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464-504.
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal, 9(2), 233-255.
Christ, O., & Schlüter, E. (2012). Strukturgleichungsmodelle mit Mplus: Eine praktische Einführung. München: De Gruyter Oldenbourg Verlag.
Correll, J., & Park, B. (2005). A model of the ingroup as a resource. Personality and Social Psychology Review, 9, 341-359.
Dyce, J. A., & Cornell, J. (1996). Factorial validity of the group environment questionnaire among musicians. The Journal of Social Psychology, 136(2), 263-264. doi:10.1080/00224545.1996.9714003
Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents’ experiences in schools and in families. American Psychologist, 48(2), 90-101.
Eder, F. (1996). Schul- und Klassenklima: Ausprägungen, Determinanten und Wirkungen des Klimas in höheren Schulen. Innsbruck: Studien-Verlag.

Estabrooks, P. A., & Carron, A. V. (2000). The physical activity group environment questionnaire: An instrument for the assessment of cohesion in exercise classes. *Group Dynamics: Theory, Research, and Practice, 4*(3), 230-243. doi:10.1037/1089-2699.4.3.230

Eys, M. A., Loughead, T. M., Bray, S. R., & Carron, A. V. (2009). Development of a cohesion questionnaire for youth: The youth sport environment questionnaire. *Journal of Sport and Exercise Psychology, 31*(2), 390-408.

Festinger, L. (1950). Informal social communication. *Psychological Review, 57*(5), 271-282.

Fraser, B. J., Anderson, G. J., & Walberg, H. J. (1982). Assessment of learning environments: Manual for learning environment inventory (LEI) and my class inventory (MCI). Perth: Western Australian Institute of Technology.

Fraser, B. J., & Fisher, D. L. (1983). Development and validation of short forms of some instruments measuring student perceptions of actual and preferred classroom learning environment. *Science Education, 67*(1), 115-131.

Fraser, B. J., McRobbie, C. J., & Giddings, G. J. (1993). Development and cross-national validation of a laboratory classroom environment instrument for senior high school science. *Science Education, 77*(1), 1-24. doi:10.1002/sce.3730770102

Geiser, C. (2011). *Datenanalyse mit Mplus: Eine anwendungsorientierte Einführung* (2. Auflage). Wiesbaden: Springer VS.

Gerber, J., & Wheeler, L. (2009). On being rejected: A meta-analysis of experimental research on rejection. *Perspectives on Psychological Science, 4*(5), 468-488. doi:10.1111/j.1745-6924.2009.01158.x

Hagenauer, G., & Hascher, T. (2011). Schulische Lernfreude in der Sekundarstufe 1 und deren Beziehung zu Kontroll- und Valenzkognitionen. *Zeitschrift für Pädagogische Psychologie, 25*(1), 63-80.

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2014). *Multivariate data analysis*. Essex: Pearson.

Hammer, S., Reiss, M., Lehner, M. C., Heine, J.-H., Sälzer, C., & Heinze, A. (2016). Mathematische Kompetenz in PISA 2015: Ergebnisse, Veränderungen und Perspektiven. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Koller (Ed.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (pp. 219-248). Münster: Waxmann.

Helmke, A. (1997). Entwicklung lern- und leistungsbezogener Motive und Einstellungen: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Ed.), *Entwicklung im Grundschulalter* (pp. 59-76). Weinheim: Beltz.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Johnson, A. L., Crawford, M. T., Sherman, S. J., Rutchick, A. M., Hamilton, D. L., Ferreira, M. B., & Petrocelli, J. V. (2006). A functional perspective on group memberships: Differential need fulfillment in a group typology. *Journal of Experimental Social Psychology, 42*(6), 707-719. doi:10.1016/j.jesp.2005.08.002

Kleinknecht, C., Kleinert, J., & Ohlert, J. (2014). Erfassung von “Kohäsion im Team von Freizeit- und Gesundheitssportgruppen” (KIT-FG). *Zeitschrift für Gesundheitspsychologie, 10*(1), 39-51.

Li, F., & Harmer, P. (1996). Confirmatory factor analysis of the group environment questionnaire with an intercollegiate sample. *Journal of Sport and Exercise Psychology, 18*(1), 49-63. doi:10.1123/jsep.18.1.49

Martin, L. J., Carron, A. V., Eys, M. A., & Loughead, T. M. (2012). Development of a cohesion inventory for children’s sport teams. *Group Dynamics: Theory, Research, and Practice, 16*(1), 68-79.

Muthén, L., & Muthén, B. (1998-2018). *Mplus* (version 7). Los Angeles, CA: Muthén & Muthén.

Ohlert, J. (2012). Kohäsionsfragebogen für Individual- und Teamsport - Leistungssport (KIT-L): A German-language instrument for measuring group cohesion in individual and team sports. *International Journal of Sport and Exercise Psychology, 10*(1), 39-51.

Oswald, F. (1989). *Schulklima: Die Wirkungen der persönlichen Beziehungen in der Schule* (1. Auflage). Wien: Universitätsverlag.

Rauch, D., Mang, J., Hüttig, H., & Haag, N. (2016). Naturwissenschaftliche Kompetenz von Schülerinnen und Schülern mit Zuwanderungsgeschichte. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, &
O. Köller (Ed.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (pp. 317-348). Münster: Waxmann.

Schiepe-Tiska, A, Rönnebeck, S., Schöps, K., Neumann, K., Schmidtner, S., Parchmann, I., & Prenzel, M. (2016). Naturwissenschaftliche Kompetenz in PISA 2015 - Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller (Ed.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (pp. 45-98). Münster: Waxmann.

Schürer, S. (2019). Der Einfluss von Gruppenkohäsion auf die soziale Partizipation individueller Schülerinnen und Schüler - Eine Interventionsstudie an inklusiv unterrichtenden Grundschulen (Dissertation). Retrieved from https://d-nb.info/1193126185/34

Schutz, R. W., Eom, H. J., Smoll, F. L., & Smith, R. E. (1994). Examination of the factorial validity of the group environment questionnaire. *Research Quarterly for Exercise and Sport, 65*(3), 226-236. doi:10.1080/02701367.1994.10607623

Van Bergen, A., & Koekebakker, J. (1959). “Group cohesiveness” in laboratory experiments. *Acta Psychologica, 16*, 81-98. doi:10.1016/0001-6918(59)90087-3

Whitton, S. M., & Fletcher, R. B. (2014). The group environment questionnaire. *Small Group Research, 45*(1), 68-88. doi:10.1177/1046496413511121

Zander, L., Kreutzmann, M., & Hannover, B. (2017). Peerbeziehungen im Klassenzimmer. *Zeitschrift für Erziehungswissenschaft, 20*(3), 353-386. doi:10.1007/s11618-017-0768-9