Social Sciences & Humanities Open 12 (2025) 101827

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

Cultural diversity beyond the binary – testing a disparity-weighted indicator of cultural classroom composition

Merle-Sophie Thielmann *, Benedict C.O.F. Fehringer, Karina Karst

School of Social Sciences, University of Mannheim, Germany

ARTICLE INFO

Keywords:
Cultural diversity
Cultural values
Classroom composition
Reading achievement
Diversity indices

ABSTRACT

Cultural classroom diversity is theorized to affect individual students through several mechanisms related to different diversity aspects. These include the number, relative size, and dissimilarity of cultural groups, not all of which are simultaneously captured by diversity measures currently established in educational contexts. Disparity-weighted measures, which quantify and incorporate differences between cultural groups, may offer a more comprehensive and valid operationalization of cultural diversity. This study adapts and tests the cultural fractionalization index *CF* from sociology, which accounts for similarities of cultural values between cultural groups. We examine *CF* in comparison to established diversity indicators using a longitudinal sample of 1560 students from 65 German secondary school classrooms. *CF* strongly correlated with established diversity indices but notably differentiated better between highly diverse classrooms, suggesting it captured additional information on cultural classroom composition. In exploratory multilevel analyses, we further compared *CF* to established measures in the common application of predicting individual students' reading achievement. Results indicated significant negative effects of most diversity measures on reading achievement but a diverging pattern of results for *CF,* possibly reflecting its ability to correct overestimations of diversity inherent in established indices by including disparity. These findings highlight the potential of disparity-weighted indices like *CF* to advance research on the effects of cultural diversity and its potential underlying mechanisms. Our results emphasize the importance of a theory-driven definition of diversity and a deliberate choice of corresponding diversity measures.

1. Introduction

As the environment students spend most of their day in, the classroom has great potential to influence their academic and social development (Dumont et al., 2013). In Germany, past and recent immigrant has shaped student populations: In 2023, 42 % of school-age children had an immigrant background, and 14 % had migrated to Germany themselves (Sachverständigenrat Integration und Migration, 2025). National and international research has explored how cultural classroom composition affects student outcomes, especially achievement (Dronkers & van der Velden, 2013; Mickelson et al., 2021). Earlier studies have commonly used the proportion of minority students to measure cultural classroom composition (van Ewijk & Sleegers, 2010a). The focus of recent research has increasingly shifted from this binary minority-majority distinction toward the diversity of cultural groups, which affects student achievement by shaping student interactions and teachers' instruction (Rjosk et al., 2017; Veerman et al., 2013). This categorial view of cultural group membership represents an important step toward a more nuanced conceptualization of cultural background, corroborated by the improved fit of models including diversity measures (Veerman et al., 2013). Yet, it does not adequately capture linguistic and value differences between cultural groups, which theory suggests are central to diversity effects.

This paper addresses this gap by using an index that not only includes group number and size, but also the dissimilarity between cultural groups: the cultural fractionalization index *CF*, previously used in linguistic and sociological research (Bredtmann et al., 2021; Greenberg, 1956; Schaeffer, 2013). We adapted *CF* to measure cultural classroom diversity, focusing on cultural values as one meaningful dimension of dissimilarity and developing a weighting procedure which reflected value differences between students. To evaluate whether this extended operationalization yielded a more valid measure of cultural diversity, we compared *CF* to established diversity measures in educational research and examined their correlations, their resulting values in

* Corresponding author. Department of Psychology, University of Mannheim, A5,6 – Building B, 68159, Mannheim, Germany.
E-mail address: thielmann@uni-mannheim.de (M.-S. Thielmann).

https://doi.org/10.1016/j.ssaho.2025.101827
Received 4 July 2024; Received in revised form 18 July 2025; Accepted 21 July 2025
Available online 28 July 2025

classrooms of different composition, and their performance in a common application: the prediction of student achievement, specifically reading achievement as a central basic competency in education (Autor:innengruppe Bildungsberichterstattung, 2022). For comparison with established findings, we tested the measures both as stand-alone predictors and when controlling for the minority proportion.

1.1. Cultural diversity in the classroom

Cultural diversity, also referred to as origin or ethnic diversity, is an aspect of classroom composition describing the heterogeneity of students' ethnic or cultural backgrounds (Budescu & Budescu, 2012). These backgrounds not only encompass their binary immigrant status, cultural origin, and migration generation, but also related characteristics like language, socioeconomic status, and values–all of which can shape classroom processes and experiences (Mickelson et al., 2021; Nemetz & Christensen, 1996; van Ewijk & Sleegers, 2010a). This complexity poses a challenge to the valid conceptualization and operationalization of cultural diversity: Researchers must specify which cultural background characteristics they aim to capture and use suitable measures to represent them and their diversity (Bredtmann et al., 2021).

1.2. Operationalization of cultural diversity within the classroom

Diversity measures should be contrasted with the commonly used proportion of cultural minority students (e.g., Mickelson et al., 2021; Mok et al., 2016), which is based on a dichotomous distinction of cultural minority and majority. This makes it inadequate for capturing diversity, which encompasses the number, size, and dissimilarity of cultural groups and thus requires differentiating between cultural minority groups (Budescu & Budescu, 2012; Schaeffer, 2016). Diversity indices differ in their ability to capture these aspects.

1.2.1. Variability

Variability describes the number of cultural groups in a classroom. The more cultural groups it contains, the more diverse a classroom. Perhaps the simplest measure of cultural diversity, solely capturing variability, is the number of categories *Ncat* (Driessen, 2002; Dotzel et al., 2021).

1.2.2. Balance

According to most common conceptualizations of diversity, variability alone insufficiently describes classroom diversity (Budescu & Budescu, 2012). The relative size of cultural groups, also referred to as balance, should also be accounted for: Classrooms with more cultural groups of relatively equal size are considered more diverse than classrooms with fewer groups or disproportionately large singular groups (Schaeffer, 2016).

Diversity indices like Shannon's H and the Simpson-Gini index D (Dotzel et al., 2021; Rjosk et al., 2017) include variability and balance in one index, although they give different weight to them (Budescu & Budescu, 2012):

$$D = 1 - \sum_{i=1}^{k} p_i^2$$

$$H = -\sum_{i=1}^{k} p_i \ln(p_i),$$

where k indicates the number of cultural groups per classroom and p_i denotes the classroom proportion of each cultural group i. Both indices are 0 if all students belong to a single cultural group and 1 if all k groups have the same size. D is especially intuitive to interpret, representing the probability that two randomly chosen students from a classroom belong to different cultural groups (Rjosk et al., 2017): The more cultural groups are present within a classroom, and the more evenly sized they are, the higher the probability that two randomly drawn students differ in cultural background, and thus the higher the diversity of this classroom.

We illustrate these aspects of diversity using the comparison of fictional classrooms A and B in Fig. 1: Both are identical in size ($n = 25$) and variability ($Ncat = 5$) but differ in balance. In classroom A, all cultural groups are evenly sized, while classroom B is dominated by one disproportionately large group. Once balance is considered, classroom A would be considered more diverse, as reflected by its higher Simpson-Gini index ($D = .80$ vs. $D = .68$).

1.2.3. Disparity

All previously discussed diversity measures employ a categorial understanding of cultural backgrounds, commonly assigned based on countries of origin or coding procedures (Statistisches Bundesamt, 2022). They distinguish only whether two groups are distinct and therefore are both included in the index, or not. By doing this, they assume all groups to be equally dissimilar. However, cultural group membership can also be associated with differences in values and other characteristics. This dissimilarity is captured by the third component of diversity, disparity (Schaeffer, 2016). The greater the dissimilarity of groups, the higher a classroom's diversity. The comparison of classrooms B and C in Fig. 1 illustrates this: Both are identical regarding the number and size of cultural groups, and therefore in variability and balance. However, students in classroom C come from geographically more similar backgrounds, which might share similar languages, histories, and cultural values. Accounting for disparity, classroom C would thus be considered less diverse than classroom B.

Measuring disparity requires moving beyond a categorial conceptualization of cultural backgrounds to include group similarity on one or more continuous variables into index calculation. In linguistic research, this has led to the development of the Greenberg index, closely related to the Simpson-Gini index D (Bredtmann et al., 2021; Greenberg, 1956). While D weights each cultural group with its own proportion, the Greenberg index considers pairs of groups: For each combination of groups i and j, their proportions p_i and p_j are multiplied by a so-called disparity weight r_{ij} reflecting their similarity.

$$GI = \sum_{i=1}^{k} \sum_{j=1}^{k} r_{ij} p_i p_j$$

These weights range from 0 for maximum dissimilarity to 1 for culturally identical groups. The inverse of GI equals D if all combinations i, j with $i \neq j$ are weighted with $r_{i,j} = 0$. This illustrates the assumption made by categorial conceptualizations of cultural background: They assume that all groups are not only equally but also maximally dissimilar (Schaeffer, 2016). As similarity between groups increases, the inverse of GI yields comparatively smaller values than D, reflecting that based on disparity, diversity is considered lower as groups become more similar.

Disparity-weighted measures provide a more valid measurement of cultural diversity in two ways: First, by capturing all three diversity aspects, they align more closely with theoretical understandings of diversity. By additionally allowing researchers to specify group differences of interest and include this information in index calculation, they also potentially allow a more accurate measurement of cultural diversity (Frongillo et al., 2019). Second, they may reduce bias by correcting the assumption of maximum dissimilarity made by other diversity measures. This simplification likely does not reflect classroom reality, which might overestimate cultural diversity in classrooms with similarities between groups (Schaeffer, 2016).

Disparity-weighted measures have been applied in several contexts related to cultural diversity, including in classroom contexts: Measures of phonetic similarity have been used to calculate disparity weights for the Greenberg index to investigate linguistic classroom diversity (Bredtmann et al., 2021). For cultural diversity, the related

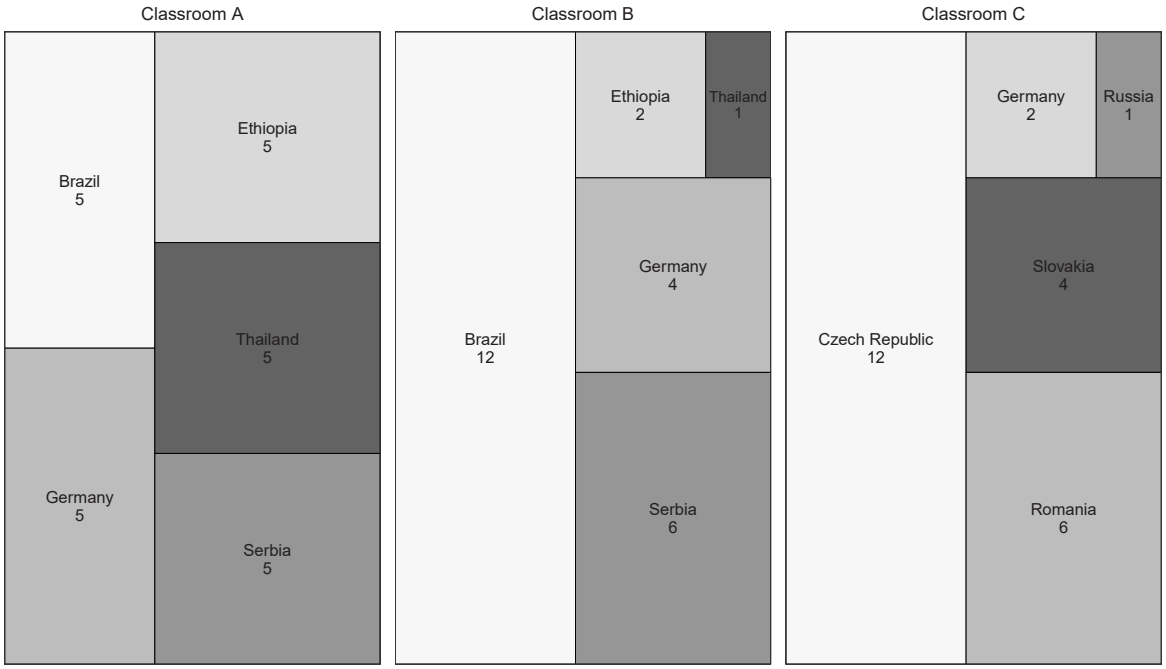

Fig. 1. Illustration of diversity components.
Note. Hypothetical classrooms of equal size ($n = 25$ students) and variability ($Ncat = 5$), differing in balance and disparity.

fractionalization index *CF* has been used in sociological studies on group cohesion and public goods provision: Differences between cultural groups have been operationalized using linguistic similarity (Baldwin & Huber, 2010), but also cultural values, measured through scales from the World Values Survey (Haerpfer et al., 2022; Inglehart & Welzel, 2005), to represent cultural differences (Schaeffer, 2013). The latter approach aligns closely with our conceptualization of cultural diversity in this study. We thus adapted this method, making adjustments to address current shortcomings and improve its suitability for classroom contexts.

1.3. Cultural diversity and student outcomes

As illustrated, cultural diversity is a multifaceted construct encompassing the number, balance, and (dis)similarity of cultural groups. As a part of cultural classroom composition, these aspects shape the learning environment and thus influence individual student outcomes, including achievement as a central educational outcome (Dumont et al., 2013). This influence can arise through links between classroom composition and school resources, teaching quality, and peer interactions (Rjosk, 2022). Theoretical considerations on cultural diversity have focused mostly on the latter two and differ in the diversity aspects that they deem relevant to these processes.

Some mechanisms focus on the effects of overall group membership, employing a categorial view of cultural background and focusing only on variability and balance. For instance, high variability has been shown to result in more negative teacher attitudes toward minority students (Glock et al., 2019), which might negatively affect instruction and thus student achievement. Variability and balance also determine the number of same-culture contacts available to each student. Such contacts can benefit students' sense of belonging, parental involvement, and achievement (Benner & Yan, 2015; Mok et al., 2016; Rjosk et al., 2017). In highly diverse classrooms, contact opportunities are more limited, which may negatively affect these outcomes. Recent studies question whether in-group representation represents an individual-level construct separate from diversity as a group-level characteristic (Chan & Benner, 2024).

Many mechanisms consider disparity as well, focusing on differences between students associated with cultural diversity that shape peer interactions and teaching practices. Linguistic diversity, especially differences in students' proficiency in the language of instruction, might affect teaching (Driessen, 2002) and student interactions (Bredtmann et al., 2021; van Ewijk & Sleegers, 2010a), but empirical findings have been mixed (Bredtmann et al., 2021; Dotzel et al., 2021; Driessen, 2002). Socioeconomic differences, often associated with cultural background, have also been shown to affect student achievement (van Ewijk & Sleegers, 2010b). Findings indicate that part of the effects of cultural diversity may be explained by socioeconomic differences, since including indicators of socioeconomic classroom composition often reduces them (De Schaepmeester et al., 2022; Dumont et al., 2013).

While cultural diversity can be associated with a range of characteristics that differentiate students with different cultural backgrounds, most theories implicitly or explicitly assume particularly differences in cultural values to underlie its effects on classroom processes and student outcomes (Braster & Dronkers, 2015; van Ewijk & Sleegers, 2010a; Veerman et al., 2013). Cultural values refer to shared patterns of beliefs and attitudes that distinguish groups and can shape social processes (Kaasa & Minkov, 2022), including classroom processes influencing student achievement. The predicted direction of effects varies: Engaging with contrasting values of peers in more diverse classrooms may benefit cognitive development and ultimately achievement (Piaget, 1977). Conversely, greater similarity in students' values might reduce conflicts and strengthen identification with and support between peers (Dronkers & van der Velden, 2013; Mok et al., 2016), enhancing their feeling of belonging, motivation and achievement (Rjosk et al., 2017; Schachner et al., 2019). Greater differences in values might also challenge teachers when adapting lessons to students' backgrounds in terms of culturally responsive teaching and managing conflicts between cultural groups (Braster & Dronkers, 2015; Gebauer & McElvany, 2020), potentially lowering teaching quality and thus achievement.

In sum, theoretical predictions on effects of the diversity of cultural values on student achievement have been mixed. Peer interactions may lead to both positive and negative effects, and negative effects may result from reduced instructional quality. Empirical findings have been

similarly inconsistent, varying in both the studied diversity aspects and achievement domains: The variety of cultural groups showed negative effects on students' language performance only before controlling for other school composition characteristics (Driessen, 2002), and none on reading or mathematics achievement (Veerman et al., 2013). Cultural diversity indices have positively predicted mathematics achievement (Rjosk et al., 2017), though one study found significant effects only for minority students (Braster & Dronkers, 2015). Diversity indices have shown inconsistent associations with reading achievement, with some studies finding significant positive effects only for minority students (Maestri, 2016), or negative effects only in higher grades (Veerman et al., 2013), while a study from the German context found no overall effect at all (Rjosk et al., 2017). In summary, empirical findings remain inconclusive and effects differ between achievement domains and sometimes student groups.

1.4. Summary and research questions

Cultural diversity is a complex construct that encompasses several aspects and relates to various differences between students potentially relevant to classroom processes and student outcomes. This complexity poses a major challenge to its valid measurement. While many theoretical predictions are based on disparity, few studies have used measures able to includes this diversity aspect (Dronkers & van der Velden, 2013; Mok et al., 2016). Disparity-weighted indices allow researchers to define and capture the relevant differences between students within a study's theoretical framework, potentially resulting in a more comprehensive and thus valid operationalization of cultural diversity. While such measures have been used for linguistic classroom diversity (Bredtmann et al., 2021), they have not yet been used to examine cultural diversity in classrooms.

This study aimed to adapt and test such a disparity-weighted index for measuring cultural diversity in classroom contexts. Specifically focusing on cultural diversity as the dissimilarity between students regarding cultural values, we used a disparity-weighted measure to operationalize these differences.

Using the publicly available World Values Survey (Haerpfer et al., 2022), we developed a procedure to calculate disparity weights representing cultural value similarities between countries. These weights were then applied to calculate the cultural fractionalization index *CF*. We compared *CF* to other diversity measures used in educational research to establish whether it captured additional information on cultural diversity, formulating the following hypotheses and research questions.

Our first objective was to compare *CF* to established diversity indices. Based on the way *CF* is calculated and operationalizes diversity, we formulated the following hypotheses:

(1) We expect positive correlations between *CF* and other established diversity indices.
  (1.1) *CF* should show the highest correlation to the Simpson-Gini index *D*.

This assumption rested on the calculation of *CF* as an extension of *D,* with added weights accounting for disparity.

  (1.2) *CF* should show the lowest correlations to indices representing single aspect of diversity, like *Ncat* and the minority proportion *pmig*.

To further examine whether *CF* offered a more comprehensive and valid measurement of cultural diversity, we evaluated its performance alongside established indices as a predictor of individual student achievement (Frongillo et al., 2019; Mickelson et al., 2021). These analyses aimed to examine *CF*'s behavior in a common use case, which allows comparison to an extensive body of findings, rather than draw definitive conclusions about effects of cultural diversity on achievement. We specifically focused on reading achievement, as several mechanisms suggest a particular relevance of cultural values, and thus possibly their diversity, to this achievement domain: Students' cultural values and experiences can shape how they engage with texts, with alignment between a texts' cultural content and students' cultural background improving comprehension, motivation, and achievement (Kamil et al., 2011; Rupley et al., 2008). Cultural backgrounds might also shape attitudes towards reading at home, e.g., the importance placed on reading or the role of parents in supporting students' literacy development (Kamil et al., 2011). Given the inconclusive empirical findings for reading achievement and the difficulty to interpret results with regard to cultural value differences specifically, we formulated no expectations on the relationship between *CF* and reading achievement and conducted explorative analyses.

Exploration 1 How does the cultural diversity of a classroom affect individual students' reading achievement depending on different operationalizations of cultural classroom composition?

Following the approach of previous studies to isolate the effect of cultural diversity beyond the minority proportion (Bredtmann et al., 2021; Rjosk et al., 2017), we also included models testing the effects of diversity measures after including *pmig.*

Exploration 2 Does the cultural diversity of a classroom, operationalized through different diversity measures, affect individual students' reading achievement beyond the effect of overall cultural classroom composition, measured by the minority proportion *pmig*?

2. Method

2.1. Design and sample

We utilized a dataset of 79 fifth-grade classrooms from four school tracks in Southern Germany (Karst et al., 2022), collected during an intervention study. The intervention, a reading strategy training, had no significant main effect (Karst et al., 2022) and was controlled for in all our analyses for this study. In addition to reading achievement tests at both measurement points (September 2018 and January 2019), students completed a questionnaire on motivational and demographic variables at the end of the term, providing all variables necessary for secondary analysis on the effect of cultural diversity.

For our study, we excluded classrooms from the lowest secondary school track, as their number ($n = 2$) was too small. Classrooms where also excluded if less than 75 % of students participated in the study, as we did not consider this level of data coverage sufficient to calculate a representative measure of classroom composition.

The final sample of this study comprised $N = 1560$ students from 65 classrooms and 25 schools, with an average of 24 participating students ($SD = 4.23$) per classroom. 75 % of students attended academic track schools, 9 % comprehensive track, and 16 % intermediate track. Fifty-six percent of students had an immigrant background, with the most common countries of origin being Turkey (11 %), Russia (6 %), and Romania (3 %). Their mean age was 10.17 years, and 51 % of students were female.

2.2. Variables

2.2.1. Reading achievement

Beginning-of-term reading achievement was measured by the *Lernstand 5,* a standardized test of reading achievement conducted at the beginning of 5th grade in Baden-Württemberg (Institut für Bildungsanalysen Baden-Württemberg, n.d.). A parallel version of this test was used at the end of the term. Ability scores were calculated from raw data using IRT models. Item parameters for the *Lernstand 5* were estimated in

a pre-study and used to calculate ability scores for both time points to ensure comparability over time. The expected a posteriori (EAP/PV) reliability was .85 at the beginning and .86 at the end of term (Karst et al., 2022). We used end-of-term reading achievement as the outcome variable in all analyses, controlling for initial achievement.

2.2.2. Student-level independent variables

We included students' gender, binarily coded as male or female, as an individual-level context variable. As an indicator of socioeconomic background, the questionnaire included a measure of cultural capital: Participants were asked to indicate the number of books they had at home on a five-point scale ranging from close to none to more than 200. The scale was accompanied by illustrations to aid students in their answers.

To capture their cultural background, students were asked to indicate whether they and their parents were each born in Germany. If a parent was born outside of Germany, students were asked to provide that parent's country of birth. We used this information to code students' cultural minority membership and cultural background for the calculation of our diversity indicators.

2.2.3. Classroom-level independent variables

On the classroom level, we controlled for school track as a dummy-coded variable using academic track, the most common school track, as the reference group. We also included a binarily coded variable indicating membership in the intervention group to account for whether they received the reading strategy training. As central predictors on the classroom level we used five indicators measuring cultural diversity. (1) The proportion of minority students *pmig* as a dichotomous measure of balance and (2) the number of cultural groups representing variability, two indices including both variability and balance, (3) the Gini-Simpson index *D* and (4) the Shannon index *H,* and a fifth indicator based on all three diversity aspects, variability, balance, and disparity, (5) the cultural fractionalization index *CF*.

Analogous to previous papers (Rjosk et al., 2017), we used students' immigrant background to calculate indices 1 to 4: Students with at least one parent born outside of Germany were classified as having an immigrant background and were assigned the country they or their parents had migrated from as their cultural background. If parents had different non-German origins, students were assigned an "undeterminable" immigrant background (see Appendix A1). Indices 1 to 4 were calculated based on this classification and the equations above.

As a disparity-weighted diversity measure, we use the cultural fractionalization index *CF*. Its interpretation aligns closely to Shannon's *H* and the Simpson-Gini index *D*: Values range from 0 (low diversity) to 1 (high diversity). To operationalize disparity, we used weights to express the similarity of cultural value between groups, building on a procedure by Schaeffer (2013), who used the World Values Survey (WVS) to derive disparity weights. Drawing on data for 64 countries from the 7th wave (2017–2022) of the WVS (Haerpfer et al., 2022), we used the 10-item subscale on socialization goals, which asked participants to assess the importance of imparting values like hard work or imagination to children. This subscale should represent a more proximal factor in children's socialization than broader cultural values and norms (Bornstein, 2010). We extracted two orthogonal factors, each based on four dichotomous items, representing the cultural values of self-directedness vs. other-directedness (S) and civility vs. practicality (C). Mean scores on both factors were calculated for each country (Fig. 2, Panel A) (Bond & Lun, 2014).

Unlike for previous indices, we did not use students' nationality (Schaeffer, 2013) or immigrant background in the calculation of *CF*. Both of these approaches have limitations: Student nationality fails to capture the likely influence of parents on students' cultural values, especially for students born in Germany to one or two non-German parents. The coded immigrant background (Statistisches Bundesamt, 2022) addresses this to some extent yet still makes potentially inaccurate assumptions about students with parents from two different countries (25 % of our sample; Table 1). This may systematically

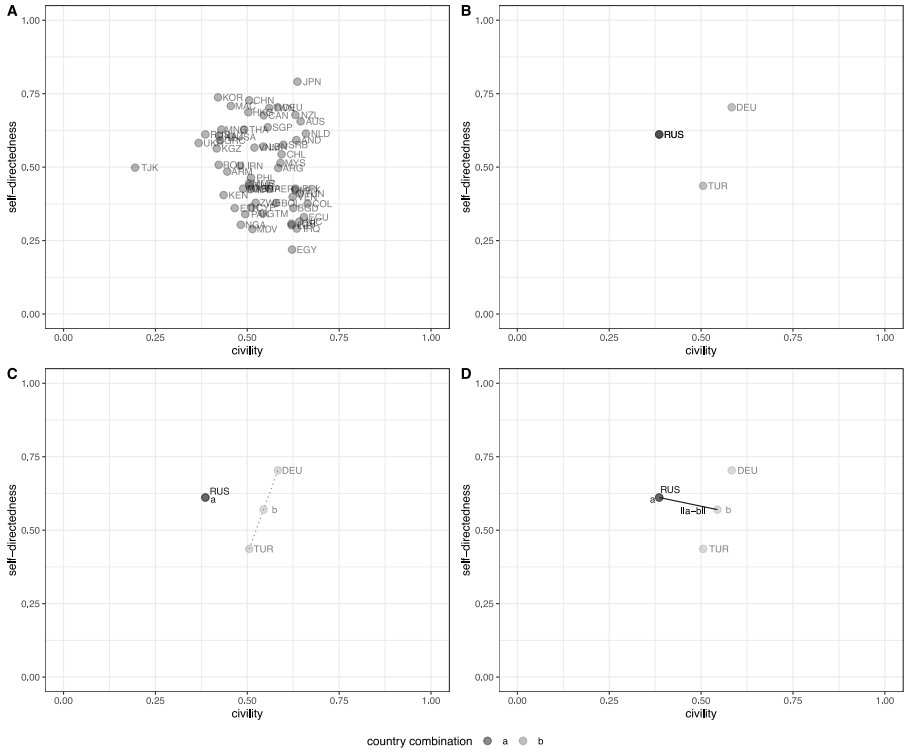

Fig. 2. Representation of student backgrounds for *CF* calculation.
Note. A: Scatterplot of cultural values from WVS countries. B: Students are represented by both their parents first. C: Individual students were assigned the average of the values of their parents' countries of origin. D: Dissimilarity was expressed as the vector distance between cultural values of respective country combinations.

Table 1
Information available for calculation of disparity weights.

Parents Identical[a]	Parental Countries of Origin Included in the WVS			
	0	1	2	*n*
0	22	176	191	389
1	155	0	930	1085
Missing	76	10	0	86
n	253	186	1121	1560

Note. $N = 1560$.
[a] 0 = Parents coming from the same country of origin, 1 = Parents coming from different countries or origin, Missing = missing information on one or both parents' countries of origin.

overestimate cultural differences in the classroom: The first overestimation concerns students with one German parent (31 % of students with an immigrant background). For example, students with one German and one Turkish parent (3 % of our sample) would be assigned a Turkish immigrant background. As a result, they would be assigned the same cultural values as students born in Turkey to two Turkish parents. However, these students' values and experiences are likely shaped by both parents' cultures, and they might also share cultural similarities with children of one or two German parents. Representing cultural background solely through immigrant background likely would have mischaracterized their cultural values and may have overestimated cultural differences in the classroom.

Furthermore, students with parents from two different non-German countries are classified as having an "undeterminable" immigrant background according to the Federal Office of Statistics, and have to be treated as a single separate category during index calculation (Rjosk et al., 2017). However, this incorrectly implies students of "undeterminable" background all have maximally similar cultural values and are maximally dissimilar from all other cultural groups, including their parents' countries of origin, further overestimating cultural differences.

To better reflect the potential influence of both parents on student values in the calculation of *CF*, we instead assigned student background based on both of their parents' birth countries. Each student was assigned the average of their parents' countries on both self-directedness (S) and civility (C). This resulted in intermediate values for students with parents from different countries.

Fig. 2 illustrates this way of representing student background using two fictional students: Student *a* has two Russian parents (represented by its ISO-code RUS), while student *b* has one parent from Germany (DEU) and one from Turkey (TUR). We matched all parents' countries of origin to numeric scores of civility (C) and self-directedness (S) from the WVS (Panel A) and represented individual students through both their parents (Panel B). Students' cultural values were calculated as the mean of their parents' countries on both C and S (Panel C). Panel C also highlights the different results of this approach for students with mixed heritage: Based on immigrant background, student *b* would have been assigned the cultural values of Turkey (TUR), which diverges considerably from the combined score based on both parents' cultural values (point *b* in Panel C).

We calculated *CF* based on the shares of every combination of parent's origin countries, with disparity weights also derived based on country combinations:

We (1) listed all possible combinations of parents' origin countries to calculate corresponding weights for each combination. We then (2) matched countries of origin to self-directedness (S) and civility (C) values from the WVS (Fig. 2, Panel B) and calculated mean S and C scores for each country combination (Fig. 2, Panel C). Each combination was assigned a vector $\vec{a}$ containing its C and S values:

$$\vec{a} = \begin{pmatrix} C_a \\ S_a \end{pmatrix} = \begin{pmatrix} \frac{(C_{a1} + C_{a2})}{2} \\ \frac{(S_{a1} + S_{a2})}{2} \end{pmatrix}$$

(3) We then compared civility and self-orientation scores of all combinations. To determine their similarity, we used the vector distance between combinations: smaller distances indicated that two combinations *a, b* were more similar in cultural values (Fig. 2, Panel D).

$$\left\| \vec{a} - \vec{b} \right\| = \sqrt{(C_a - C_b)^2 + (S_a - S_b)^2}$$

(4) Since disparity weights needed to range from 0 for maximum dissimilarity and 1 for maximum similarity, we standardized the resulting distances in an extension of the approach by Schaeffer (2013). We divided each distance by the maximum possible distance, which would result from comparing two country combinations with the smallest (0) and largest (1) respective values of C and S $\left\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\| = \sqrt{(1-0)^2 + (0-1)^2} = \sqrt{2}$.

$$r_{ab} = 1 - \frac{\left\| \vec{a} - \vec{b} \right\|}{\sqrt{2}} = 1 - \frac{\sqrt{(C_a - C_b)^2 + (S_a - S_b)^2}}{\sqrt{2}}$$

(5) We then used these disparity weights to calculate the final index based on country combinations *a*, *b*:

$$CF = 1 - \sum_{a=1}^{n} \sum_{b=1}^{n} r_{ab} p_a p_b$$

Students in our sample came from 79 different countries, not all of which were covered by the WVS. Table 1 shows the information available for the calculation of disparity weights. For 1121 students (72 % of the sample), both parents came from countries included in the WVS, meaning all necessary information for weight calculation was available. For 177 students (11 %), neither parents' countries were covered by the WVS, and 5 % had missing values on one or more origin variables, thus missing necessary information to calculate disparity weights. These students were conservatively assumed to be maximally dissimilar from other students in analogy to the assumption underlying categorial indices like *D*. This left 186 students with WVS data available for only one parent. To include this additional 12 % of students in the calculation of *CF*, we developed a conservative estimation procedure for their possible cultural values. This procedure is detailed in Appendix A3, with additional robustness checks reported in Appendix A4.

2.3. Analyses

All analyses were conducted using R (R Core Team, 2020). Analysis code is available upon request from the corresponding author.

2.3.1. Comparison of diversity indices

We descriptively compared distributions and scatterplots of all diversity indices, examining relationships between them, especially correlations of *CF* to established indicators of cultural diversity, to address our first study objective and hypotheses.

2.3.2. Prediction of student achievement

To address our explorative research question, we calculated multilevel models using the lme4-package (Bates et al., 2015). For each diversity index, we estimated random intercept linear regression models with fixed slopes across classrooms. Each model used one diversity index to predict end-of-term reading achievement at the individual level. Control variables were school track on a classroom level and gender,

cultural capital, intervention group, beginning-of-term reading achievement, and cultural background on an individual level.

For all diversity indices, we calculated two models, (1) one using only the respective index as the classroom-level predictor and (2) one additionally controlling for minority proportion. This followed the approach of prior studies to test the indices' predictive potential beyond the minority proportion. Effects of cultural diversity were modelled as so-called composition effects, controlling for individual student backgrounds. Student background was operationalized to closely match the way it was represented in each respective diversity index: binarily coded (0 = no immigrant background, 1 = immigrant background) for the minority proportion, as dummy variables coding the five largest origin groups for *Ncat*, *D,* and *H,* and as values of both parents on civility and self-directedness for *CF*.

Students' beginning-of-term reading achievement was standardized at the group mean to account for differences in average classroom achievement (Enders & Tofighi, 2007). To ensure comparability across classrooms with different numbers of cultural groups, we normalized *D* and *H* (Budescu & Budescu, 2012). *Ncat* was also normalized by dividing by total classroom size. All diversity indices as well as cultural capital were additionally centered at the grand mean. Categorial variables like gender and intervention group were not standardized.

2.3.3. Treatment of missing data

Of the 1560 students included in our final dataset, 12.4 % had missing values on the dependent variable, end-of-term reading achievement. We addressed this by multiple imputation using the package mice (van Buuren & Groothuis-Oudshoorn, 2011). Predictors in the imputation included student demographic variables (except cultural background), beginning-of-term reading achievement, and questionnaire items on students' ability-related self-concept and enthusiasm for reading at the beginning of term. We imputed 60 datasets with 30 iterations each (van Buuren & Groothuis-Oudshoorn, 2011).

3. Results

Descriptive statistics for all diversity measures are displayed in Table 2. Classrooms in our sample contained $3 \leq Ncat \leq 14$ countries of origin ($M = 8.57$, $SD = 2.14$). On average, 56 % of students had an immigrant background according to the Federal Office of Statistics ($SD = .17$). Of all diversity indices, *D* showed the highest values ($M = .86$), indicating that classrooms had a high variability and relatively balanced group sizes. Values of *H* were similarly high, and close to the upper limit of 1. The comparatively lower average of *CF* indicates that cultural groups were at least somewhat similar rather than maximally dissimilar.

3.1. Descriptive comparison of diversity indices

To address the first goal of our study, comparing *CF* to established indicators of cultural diversity in the classroom, and to address the corresponding hypotheses, we examined the distributions and correlations of all diversity indicators. Fig. 3 depicts scatterplots and pairwise correlations of all calculated diversity measures. A full correlation table of all variables can be found in Appendix A2. The established diversity measures (*pmig, Ncat, D,* and *H*) showed high correlations among each other, $.41 \leq r \leq .82$, $p < .001$. Both *D* ($r = .74$, $p < .001$) and *H* ($r = .94$, $p < .001$) also showed large and significant correlations to the proportion of minority students within a classroom.

CF showed significant positive correlations to all other diversity indicators ($.17 \leq r \leq .79$, $p > .001$). The association between *CF* and *D* was expectedly large due to their shared mathematical origin ($r = .79$, $p < .001$), but classrooms' diversity according to *CF* tended to be lower than that according to *D*. *CF* showed lower correlations to single-aspect measures of diversity *Ncat* ($r = .17$, $p < .001$) and *pmig* ($r = .44$, $p < .001$) than other diversity indices. The correlation between *CF* and *pmig* was significantly smaller than that of *H* ($z = -40.56$, $p < .001$) and *D* ($z = -23.85$, $p < .001$) to *pmig* (Meng et al., 1992). The same was true for the correlation between *CF* and *Ncat*. Scatterplots illustrate this as well: Classrooms with a high proportion of minority students tended to also show higher values of *D* and *H*. This close relationship only dissolved in certain value ranges of *pmig*. However, classrooms across the whole range of minority proportions still differed considerably in their diversity when disparity was also accounted for by using *CF*.

Table 2
Descriptive statistics of diversity measures.

Diversity index	*M*	*SD*	Min	Max	Range	Skewness	Kurtosis
pmig	.56	.17	.25	.95	.70	.07	−.61
Ncat	8.57	2.14	3	14	11	.12	.09
Simpson's *D*[a]	.76	.10	.44	.90	.46	−1.27	1.19
Shannon's *H*[a]	.78	.10	.5	.96	.46	−.72	−.17
CF	.44	.14	.10	.80	.70	−.08	−.01

Note. $N = 65$.
[a] Standardized according to Budescu and Budescu (2012).

3.2. Effects of cultural diversity on reading achievement

To compare *CF* to other diversity indices in a commonly used application, we conducted multilevel regression analyses using different diversity indices to predict individual students' end-of-term reading achievement. First, a model without any predictors was run to estimate the intraclass correlation coefficient (ICC). Our empty model resulted in an ICC = .22, indicating that 22 % of the variance in reading achievement could be attributed to differences between classrooms. We calculated separate random intercept models for each diversity indicator, as shown in Table 3.

Across all models, individual and classroom-level control variables showed the same effects. We illustrate them using coefficients from Model 1 (which used *pmig* as a diversity indicator). Higher beginning-of-term reading achievement predicted a higher end-of-term achievement ($b = .60$, $p < .001$), as did a higher cultural capital ($b = .06$, $p < .05$) and female gender ($b = .23$, $p < .001$). Notably, immigrant background was no significant predictor of achievement (see Models 1 and 2). Neither were individual students' cultural backgrounds nor the cultural values associated with their parents' cultures of origin (see Models 3 to 5), with one exception: Turkish students showed significantly lower reading achievement than German students ($b = -.18$, $p < .05$). On a school level, students from non-academic tracks showed significantly lower reading achievement ($-1.10 < b < -.94$, $p < .001$). The reading strategy intervention did not significantly predict student achievement in any of the models.

The minority proportion *pmig* significantly predicted reading achievement at the end of the term ($b = -.18$, $p < .001$). The other established diversity indices ($-.16 < b < -.09$, $p < .05$) also showed negative effects on student achievement when included as sole predictors (see Models 2 to 4). *CF* showed a diverging pattern of results: It had no significant effect on reading achievement ($b = -.07$). Explained variance was similar across models ($.47 < R^2 < .49$).

We calculated a second set of models (see Table 4) to investigate the effects of diversity on reading achievement after controlling for *pmig*.[1] None of the diversity indicators showed significant effects beyond *pmig* ($-.03 \leq b \leq .05$), which remained a significant predictor of student achievement in all models. Including both predictors explained a slightly higher proportion of variance compared to previous models 48 $\leq R^2 \leq .49$).

[1] It should be noted that the correlation between *H* and *pmig* was very high ($r = .94$, $p < .001$), meaning the inclusion of both of these predictors could have presented issues of collinearity. We refrain from interpreting this model and base our comparison of *CF* to established indices on the results of other models.

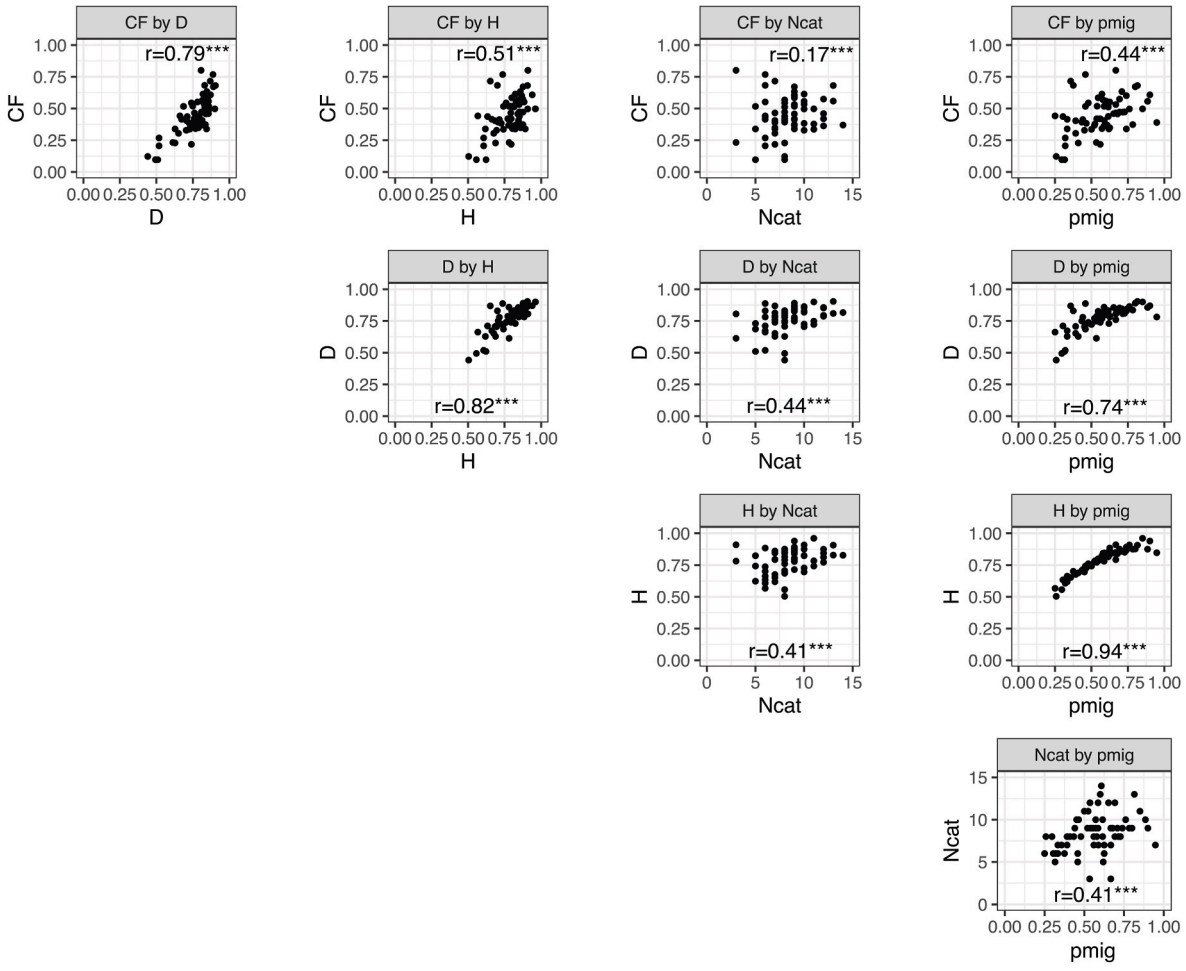

Fig. 3. Scatterplots and correlations of cultural diversity measures.
Note. CF = Cultural fractionalization index, *Ncat* = Number of cultural groups, *pmig* = Cultural minority proportion, *D* = Simpson-Gini index *D, H* = Shannon's *H*, standardized according to Budescu and Budescu (2012).
*** $p < .001$.

4. Discussion

The goal of this study was to adapt a disparity-weighted index for the measurement of cultural diversity, which we defined as the diversity of cultural values within a classroom. Such indices offer the advantage of mathematically capturing not just the number and relative size of cultural groups but also their similarity and should thus provide a more comprehensive and valid operationalization of cultural diversity as a theoretical construct. We calculated the cultural fractionalization index *CF*, previously established in sociology, and developed a weighting procedure including cultural similarities between students which also accounted for socialization effects. To assess whether *CF* provided a more valid measurement of cultural diversity, we compared it to established diversity indices to examine whether it captured additional information on classroom composition and whether this additional information contributed to predicting reading achievement.

CF showed positive correlations to all established diversity measures, particularly *H* and *D*. *CF* and *D* showed the highest correlation, as expected due to their mathematical similarity, confirming our first hypothesis.

Descriptively, *CF* showed a much larger range (.70) than other indices (see Table 2). In our diverse sample with an average minority proportion of 56 % ($SD = 17$ %), both *H* and *D* approached the upper end of their possible value range, indicating a ceiling effect. Scatterplots indicated that *CF* displayed a comparatively larger variation across the range of *pmig* and significantly smaller associations to single-aspect diversity measures *Ncat* and *pmig* compared to other indices. This supports our hypotheses and indicates that *CF* can differentiate between diverse classrooms with high minority proportions in a way other established diversity indicators cannot.

We interpret these results as evidence that *CF* captured additional information on cultural classroom composition: Classrooms that were highly diverse based on variability and balance might still have differed in their disparity. Including this information likely allowed *CF* to still differentiate between them. *CF* also corrected for a limitation of other indices: Their categorical understanding of cultural background assumes all cultural groups to be equally and maximally dissimilar, overestimating cultural differences, as reflected by the ceiling effects of *D* and *H*. The larger range of *CF* suggests it was less affected by this issue, making it especially useful in highly diverse contexts. The more nuanced coding of cultural backgrounds likely also contributed to this more accurate estimation of cultural diversity by accounting for the influence of both parents on students' cultural values.

In our exploratory analyses we compared *CF* to established diversity indices regarding their widely studied relation to student achievement. Consistent with theoretical expectations and previous empirical findings (Mickelson et al., 2021; Mok et al., 2016), *pmig* significantly negatively predicted individual students' reading achievement. When included as sole measures of classroom composition, diversity indices also negatively predicted individual student achievement, the exception being *CF*,

Table 3
Random intercept models predicting individual students' end-of-term reading achievement.

Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	.41***	.06	.44***	.07	.42***	.06	.41***	.06	.08	.25
Student-level variables										
Reading achievement t_1	.60***	.02	.59***	.02	.59***	.03	.59***	.03	.59***	.03
Gender[a]	.23***	.04	.23***	.04	.22***	.04	.23***	.04	.21***	.05
Cultural capital	.06*	.02	.06**	.02	.05*	.02	.05*	.02	.07*	.03
Immigrant background[b]	−.03	.05	−.05	.05						
Country of origin[c]										
Turkey					−.19*	.08	−.18*	.08		
Russia					−.07	.10	−.06	.10		
Romania					.05	.13	.06	.13		
Poland					.21	.15	.23	.15		
Italy					.22	.14	.21	.14		
Others					−.06	.05	−.05	.05		
Parents' cultural values[d]										
C_{mother}									−.11	.68
C_{father}									−.14	.73
S_{mother}									.07	.42
S_{father}									.65	.42
Classroom-level variables										
Intervention condition[e]	−.11	.07	−.18*	.08	−.15*	.08	−.13	.08	−.17*	.08
Intermediate track	−.94***	.10	−.98***	.11	−.87***	.10	−.90***	.10	−.88***	.11
Comprehensive track	−1.10***	.12	−1.04***	.14	−1.06***	.13	−1.07***	.13	−1.11***	.15
Pmig	−.18***	.04								
Ncat			−.09*	.04						
Simpson's *D*					−.14**	.04				
Shannon's *H*							−.16***	.04		
CF									−.07	.04
R^2	.49		.47		.48		.48		.47	

Note. $N = 1560$. The proportion of explained variance R^2 was calculated according to Snijders and Bosker (2012) using the mitml package (Grund et al., 2021).
*$p < .05$.
**$p < .01$.
***$p < .001$.
[a] 0 = Male, 1 = Female.
[b] 0 = No immigrant background, 1 = Immigrant background.
[c] Coded according to the German Federal Office of Statistics.
[d] Matched from the WVS.
[e] 0 = Control group, 1 = Participation in the reading strategy intervention.

which showed no significant effect.

This comparatively smaller, non-significant effect of *CF* might indeed represent a more accurate estimation of the effect of cultural diversity compared to the overestimation we assume to be inherent to other indices. However, established indices and *CF* describe different theoretical constructs: Established indices capture the variety and balance of cultural groups but do not further distinguish between differences in language, values, socioeconomic status, or discrimination experiences, which are associated with cultural group membership. Contrastingly, disparity-weighted measures require clear specification of and focus on specific differences of interest.

In this study, we focused on cultural value differences and operationalized them using *CF*. By comparing the effects of *CF* and established indices, our exploratory analyses also reflect the question of whether these value differences drive the effect of diversity on achievement – as hypothesized through beneficial cognitive conflicts (Piaget, 1977), interactions between students (Rjosk et al., 2017), and their relevance to instruction (Gebauer & McElvany, 2020). Our findings regarding *CF* suggest this was not the case in our study. Other factors like language differences (Bredtmann et al., 2021; Dotzel et al., 2021) or socioeconomic differences (Cascella, 2020) might be more salient in everyday classroom situations and thus more prominently shape student interactions and instruction.

Results of our models controlling for minority proportion corroborate this: Diversity indices showed no significant effect on student achievement beyond the minority proportion–consistent with other German studies (Rjosk et al., 2017). Immigrant background remains strongly associated with lower socioeconomic status and achievement results in Germany (Sachverständigenrat Integration und Migration, 2025), as reflected in our sample.

Much of the variance captured by cultural diversity was also explained by the minority proportion, indicating that students' immigrant background and associated socioeconomic factors (Sachverständigenrat Integration und Migration, 2025) may be more relevant to student achievement than cultural diversity. Nonetheless, *CF* seemed to capture additional information on cultural classroom composition, providing a more nuanced measurement of diversity useful in highly diverse contexts.

4.1. Limitations

In interpreting our results some limitations must be considered. Our data reflected a sample of highly diverse, primarily urban classrooms with a proportion of minority students above the national average. While this provided an opportunity to test *CF*'s performance in a highly diverse setting, future studies should also examine *CF* in more homogenous samples to confirm its validity across contexts.

Our data was limited to 5th grade students from mainly academic track secondary schools in Southern Germany, and to the outcome of reading achievement. While our results align with findings from representative large-scale assessments in the German context (Rjosk et al., 2017), generalizability to untracked contexts, both in other countries and school forms, and other outcomes like social and motivational outcomes may be limited.

Thirdly, the nature of our study as a secondary data analysis of an intervention study presented several limitations. The original study

Table 4
Random intercept models predicting individual students' end-of-term reading achievement controlling for minority proportion.

Variable	Model 2a		Model 3a		Model 4a		Model 5a	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	.41***	.06	.41***	.06	.42***	.06	.24	.25
Student-level variables								
Reading achievement t_1	.60***	.02	.59***	.03	.59***	.03	.59***	.03
Gender[a]	.23***	.04	.23***	.04	.23***	.04	.21***	.05
Cultural capital	.06*	.02	.05*	.02	.05*	.02	.07*	.03
Immigrant background[b]	−.03	.05						
Country of origin[c]								
Turkey			−.17*	.08	−.17*	.08		
Russia			−.05	.10	−.05	.10		
Romania			.07	.13	.07	.13		
Poland			.23	.15	.23	.15		
Italy			.22	.14	.22	.14		
Others			−.05	.05	−.05	.05		
Parents' cultural values[d]								
C_{mother}							−.19	.68
C_{father}							−.19	.73
S_{mother}							.03	.41
S_{father}							.54	.42
Classroom-level variables								
Intervention condition[e]	−.11	.07	−.11	.07	−.11	.07	−.09	.08
Intermediate track	−.93***	.10	−.92***	.10	−.94***	.10	−.92***	.10
Comprehensive track	−1.10***	.13	−1.11***	.12	−1.13***	.13	−1.18***	.13
Pmig	−.18***	.04	−.16*	.05	−.23*	.10	−.17***	.04
Ncat	−.00	.04						
Simpson's *D*			−.03	.05				
Shannon's *H*					.05	.10		
CF							.00	.04
R^2	.49		.49		.49		.48	

Note. $N = 1560$. The proportion of explained variance R^2 was calculated according to Snijders and Bosker (2012) using the mitml package (Grund et al., 2021).
*$p < .05$.
**$p < .01$.
***$p < .001$.
[a] 0 = Male, 1 = Female.
[b] 0 = No immigrant background, 1 = Immigrant background.
[c] Coded according to the German Federal Office of Statistics.
[d] Matched from the WVS.
[e] 0 = Control group, 1 = Participation in the reading strategy intervention.

evaluated a reading strategy intervention utilizing differentiated instruction. Although this intervention had no significant main effect (Karst et al., 2022) and was controlled for in all models, it may still have influenced classroom processes relevant to effects of cultural diversity. The variables available for analysis were also limited: Cultural capital, represented by the number of books at home, was used instead of a more comprehensive measure of socioeconomic status. Due to its ordinal scaling, we could not control for this variable at the classroom level. Thus, the effects of the minority proportion were likely confounded with socioeconomic disadvantages associated with both individual immigrant background and classroom-level minority proportion (Sachverständigenrat Integration und Migration, 2025; van Ewijk & Sleegers, 2010a, 2010b), possibly contributing to an overestimation of effects. Our findings are also restricted to the outcome of reading achievement. While this is a key basic competency relevant to educational and life outcomes (Autor:innengruppe Bildungsberichterstattung, 2022), including other relevant outcomes like motivation, well-being, or social integration, would have been desirable (Bredtmann et al., 2021; Rjosk et al., 2017). Future research should investigate effects of cultural diversity across a range of outcomes and test their potential interrelations in a theory-driven way.

Lastly, our conceptualization of cultural value differences based on the WVS may have resulted in limitations. While we used the Socialization Goals subscale, assuming it more closely reflected cultural differences relevant to classroom processes, this may not have accurately reflected student values. We also could not fully account for students' immigrant history, which likely also influenced their values. Future studies should directly assess students' cultural values to avoid such limitations.

4.2. Implications and conclusion

By testing an index capturing all three aspects of diversity, our study presents a step toward a more comprehensive and valid measurement of cultural diversity: As a disparity-weighted index including value differences between cultural groups, *CF* successfully captured additional information on cultural classroom composition. Both the weighting procedure and more nuanced operationalization of student background contributed to reducing the overestimation of cultural differences associated with categorial diversity measures and allowed *CF* to better differentiate between highly diverse classrooms. In such highly diverse settings, especially if many students have mixed cultural heritage, *CF* might reflect classroom composition more accurately.

Our exploratory analyses comparing *CF* and other indices as predictors of reading achievement highlight the complexity of classroom diversity: Non-significant effects of *CF* suggest that instead of cultural values, it might be differences in more salient attributes associated with cultural group membership–like linguistic backgrounds (Bredtmann et al., 2021; Dotzel et al., 2021) or socioeconomic status (Cascella, 2020)–that underlie the negative effects of diversity on reading achievement.

Disparity-weighted measures may advance research on cultural diversity both methodologically and theoretically. Recent contributions have urged researchers to explicitly define their conceptualization of diversity, select adequate measures, and examine whether these perform adequately within their study context (Chan & Benner, 2024).

Disparity-weighted indicators support this by capturing additional nuances of diversity. They may also theoretically advance the field: E.g., researchers might use the Simpson-Gini index to study group formation due to overall group membership, or disparity-weighted indices with purposeful weighting procedures (see Bredtmann et al. (2021) for linguistic differences, Schaeffer (2013) for income inequality) to identify relevant differences underlying effects of cultural diversity. Findings may inform strategies to mitigate negative impacts of diversity and support teachers in utilizing its positive potential in their classrooms.

For practical decision-making, we encourage stakeholders to be mindful of the complexity of diversity as a construct and the implications of different operationalizations. Appropriate indicators should be chosen for statistical reports and as a basis for decisions such as resource allocation. The same is true for stakeholders designing interventions to improve student outcomes in highly diverse contexts: Our results highlight the importance of considering various characteristics on which students from different cultural groups might differ, and to focus on those most relevant to the outcome of interest. Disparity-weighted indices may help them both during the initial identification of relevant differences, as well as the evaluation of interventions addressing specific differences between students.

CRediT authorship contribution statement

Merle-Sophie Thielmann: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Benedict C.O.F. Fehringer:** Writing – review & editing, Formal analysis, Data curation. **Karina Karst:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Ethical statement

The findings presented in the manuscript titled "Cultural diversity beyond the binary – testing a disparity-weighted indicator of cultural classroom composition" are the results of a secondary analysis of data originally obtained during another study. No additional data was collected specifically for this study, including none to which ethical approval applies. We do however declare that during the original collection of the data, all principles of the Declaration of Helsinki were followed. Informed consent was obtained from participants/their legal guardians.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used the tools ChatGPT and DeepL in order to help improve the clarity, conciseness, and tone of the manuscript. After using these tools, the authors reviewed and edited all generated content as needed and take full responsibility for the content of the published article. Prior to submission, the article was also reviewed and edited by a professional English language editor to ensure linguistic accuracy.

Funding

This research was supported by grants from the University of Mannheim and the Ministry of Science, Research, and the Arts in Baden-Württemberg, Germany.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssaho.2025.101827.

Data availability

The data required to reproduce the above findings cannot be publicly shared since it includes personal information about underage participants. At the time of data collection, no informed consent for public data sharing was obtained from students' legal guardians. Thus, the data cannot be shared for both legal and ethical reasons.

References

Autor:innengruppe Bildungsberichterstattung. (2022). Bildung in Deutschland 2022—Ein indikatorengestützter Bericht mit einer Analyse zum Bildungspersonal [Education in Germany - An indicator-based report including an analysis of educational staff]. *Wbv Publikation*. https://doi.org/10.3278/6001820hw

Baldwin, K., & Huber, J. D. (2010). Economic versus cultural differences: Forms of ethnic diversity and public goods provision. *American Political Science Review, 104*(4), 644–662. https://doi.org/10.1017/S0003055410000419

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benner, A. D., & Yan, N. (2015). Classroom race/ethnic composition, family-school connections, and the transition to school. *Applied Developmental Science, 19*(3), 127–138. https://doi.org/10.1080/10888691.2014.983028

Bond, M. H., & Lun, V. M. C. (2014). Citizen-making: The role of national goals for socializing children. *Social Science Research, 44*, 75–85. https://doi.org/10.1016/j.ssresearch.2013.11.002

Bornstein, M. H. (Ed.). (2010). *Handbook of cultural developmental science* (1st ed.). Psychology Press.

Braster, S., & Dronkers, J. (2015). The positive effects of ethnic diversity in classrooms on the educational performance of pupils in a multi-ethnic European metropole. *Education and Society, 33*(2), 25–49. https://doi.org/10.7459/es/33.2.03

Bredtmann, J., Otten, S., & Vonnahme, C. (2021). Linguistic diversity in the classroom, student achievement, and social integration. *Education Economics, 29*(2), 121–142. https://doi.org/10.1080/09645292.2020.1866499

Budescu, D. V., & Budescu, M. (2012). How to measure diversity when you must. *Psychological Methods, 17*(2), 215–227. https://doi.org/10.1037/a0027129

Cascella, C. (2020). Intersectional effects of socioeconomic status, phase and gender on mathematics achievement. *Educational Studies, 46*(4), 476–496. https://doi.org/10.1080/03055698.2019.1614432

Chan, M., & Benner, A. D. (2024). A consideration of racial/ethnic diversity conceptualization and measurement: Clarifying ambiguities and advancing scholarship. *Educational Psychologist, 60*(2), 69–87. https://doi.org/10.1080/00461520.2024.2376020

De Schaepmeester, L., Van Braak, J., & Aesaert, K. (2022). Social citizenship competences at the end of primary school: The role of socio-ethnic classroom diversity and teachers' citizenship beliefs and practices in the classroom climate. *Theory & Research in Social Education, 50*(2), 297–326. https://doi.org/10.1080/00933104.2022.2051106

Dotzel, S., Bonefeld, M., & Karst, K. (2021). The impact of linguistic diversity on students' reading comprehension using different diversity measures and its' differential effects depending on students' migration background. *International Journal of Educational Research, 109*, Article 101796. https://doi.org/10.1016/j.ijer.2021.101796

Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies In Educational Evaluation, 28*(4), 347–368. https://doi.org/10.1016/S0191-491X(02)00043-3

Dronkers, J., & van der Velden, R. (2013). Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using PISA data. In M. Windzio (Ed.), *Integration and inequality in educational institutions* (pp. 71–98). Springer. https://doi.org/10.1007/978-94-007-6119-3_4.

Dumont, H., Neumann, M., Maaz, K., & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen [School composition as an influencing factor on student achievement]. *Psychologie in Erziehung und Unterricht, 60*(3), 163–183. https://doi.org/10.2378/peu2013.art14d

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Frongillo, E. A., Baranowski, T., Subar, A. F., Tooze, J. A., & Kirkpatrick, S. I. (2019). Establishing validity and cross-context equivalence of measures and indicators. *Journal of the Academy of Nutrition and Dietetics, 119*(11), 1817–1830. https://doi.org/10.1016/j.jand.2018.09.005

Gebauer, M. M., & McElvany, N. (2020). Attitudes towards culturally-ethnically heterogeneous groups of students and its meaning for individualized differentiated instruction in classroom context. *Zeitschrift für Erziehungswissenschaft, 23*(4), 685–708. https://doi.org/10.1007/s11618-020-00956-8

Glock, S., Kovacs, C., & Pit-ten Cate, I. (2019). Teachers' attitudes towards ethnic minority students: Effects of schools' cultural diversity. *British Journal of Educational Psychology, 89*(4), 616–634. https://doi.org/10.1111/bjep.12248

Greenberg, J. (1956). The measurement of linguistic diversity. *Language, 32*, 109–115.

Grund, S., Robitzsch, A., & Lüdtke, O. (2021). Mitml: Tools for multiple imputation in multilevel modeling [Computer software] Version 0.4–3. https://CRAN.R-project.org/package=mitml.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kilzilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., & Puranen, B. (2022). World values survey: Round seven–country-pooled datafile version 5.0. *JD Systems Institute & WVSA Secretariat*. https://doi.org/10.14281/18241.20

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790881

Institut für Bildungsanalysen Baden-Württemberg (n.d.). Lernstand 5. Retrieved July 18th, 2025, from https://ibbw-bw.de/,Lde/Startseite/Kompetenzmessung/Lernstand+5.

Kaasa, A., & Minkov, M. (2022). Are different two-dimensional models of culture just a matter of different rotations? Evidence from the analysis based on the WVS/EVS. *Journal of Cross-Cultural Psychology, 53*(2), 127–156. https://doi.org/10.1177/00220221211072790

Kamil, M. L., Pearson, P. D., Birr Moje, E., & Afflerbach, P. P. (Eds.). *Handbook of reading research* (1st ed., Vol. IV). Routledge. https://doi.org/10.4324/9780203840412.

Karst, K., Bonefeld, M., Dotzel, S., Fehringer, B. C. O. F., & Steinwascher, M. (2022). Data-based differentiated instruction: The impact of standardized assessment and aligned teaching material on students' reading comprehension. *Learning and Instruction, 79*, Article 101597. https://doi.org/10.1016/j.learninstruc.2022.101597

Maestri, V. (2016). Can ethnic diversity have a positive effect on school achievement? *Education Economics, 25*(3), 290–303. https://doi.org/10.1080/09645292.2016.1238879

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172

Mickelson, R. A., Bottia, M. C., & Larimore, S. (2021). A metaregression analysis of the effects of school racial and ethnic composition on K–12 reading, language arts, and English outcomes. *Sociology of Race and Ethnicity, 7*(3), 401–419. https://doi.org/10.1177/2332649220942265

Mok, S. Y., Martiny, S. E., Gleibs, I. H., Keller, M. M., & Froehlich, L. (2016). The relationship between ethnic classroom composition and turkish-origin and German students' reading performance and sense of belonging. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01071

Nemetz, P. L., & Christensen, S. L. (1996). The challenge of cultural diversity: Harnessing a diversity of views to understand multiculturalism. *Academy of Management Review, 21*(2), 434–462. https://doi.org/10.2307/258668

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Viking.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing [Computer software] https://www.R-project.org/.

Rjosk, C. (2022). Dispersion of student achievement and classroom composition. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International handbook of comparative large-scale studies in education* (pp. 1–33). Cham: Springer. https://doi.org/10.1007/978-3-030-38298-8_47-1.

Rjosk, C., Richter, D., Lüdtke, O., & Eccles, J. S. (2017). Ethnic composition and heterogeneity in the classroom: Their measurement and relationship with student outcomes. *Journal of Educational Psychology, 109*(9), 1188–1204. https://doi.org/10.1037/edu0000185.supp

Rupley, W. H., Nichols, W. D., & Blair, T. R. (2008). Language and culture in literacy instruction: Where have they gone? *The Teacher Educator, 43*(3), 238–248. https://doi.org/10.1080/08878730802055180

Sachverständigenrat Integration und Migration. (2025). Ungleiche Bildungschancen—Fakten zur Benachteiligung von jungen Menschen mit Migrationshintergrund im deutschen Bildungssystem [Educational Inequalities–Facts on the deprivation of young people with an immigrant background in the German education system]. https://www.svr-migration.de/wp-content/uploads/2025/01/SVR-Fakten-zu-ungleichen-Bildungschancen-2025.pdf.

Schachner, M. K., Schwarzenthal, M., van de Vijver, F. J. R., & Noack, P. (2019). How all students can belong and achieve: Effects of the cultural diversity climate amongst students of immigrant and nonimmigrant background in Germany. *Journal of Educational Psychology, 111*(4), 703–716. https://doi.org/10.1037/edu0000303

Schaeffer, M. (2013). Can competing diversity indices inform us about why ethnic diversity erodes social cohesion? A test of five diversity indices in Germany. *Social Science Research, 42*(3), 755–774. https://doi.org/10.1016/j.ssresearch.2012.12.018

Schaeffer, M. (2016). Diversity erfassen: Statistische Diversitätsindizes [assessing diversity: Statistical diversity indices]. In P. Genkova, & T. Ringeisen (Eds.), *Handbuch diversity Kompetenz* (pp. 47–60). Springer. https://doi.org/10.1007/978-3-658-08594-0_7.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE.

Statistisches Bundesamt. (2022). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund—Ergebnisse des Mikrozensus 2021 [Population and employment. Population with an immigration background – results of the German micro-census 2021]*. German Federal Office of Statistics. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/migrationshintergrund-2010220217004.pdf?_blob=publicationFile.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1–67. https://doi.org/10.18637/jss.v045.i03

van Ewijk, R., & Sleegers, P. (2010a). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement, 21*(3), 237–265. https://doi.org/10.1080/09243451003612671

van Ewijk, R., & Sleegers, P. (2010b). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*(2), 134–150. https://doi.org/10.1016/j.edurev.2010.02.001

Veerman, G. J. M., van de Werfhorst, H. G., & Dronkers, J. (2013). Ethnic composition of the class and educational performance in primary education in the Netherlands. *Educational Research and Evaluation, 19*(5), 370–401. https://doi.org/10.1080/13803611.2013.788851