



OPEN ACCESS

EDITED BY

Gregor Ross Maxwell,
UiT The Arctic University of Norway, Norway

REVIEWED BY

Roald Øien,
UiT The Arctic University of Norway, Norway
June Forsberg,
UiT The Arctic University of Norway, Norway

*CORRESPONDENCE

Christiane Lange-Kuettner
✉ c.langekuettner@uni-bremen.de

RECEIVED 21 May 2024

ACCEPTED 02 September 2024

PUBLISHED 03 October 2024

CITATION

Lange-Kuettner C (2024) Are school grades correlated with competencies in secondary school pupils with special needs? *Front. Educ.* 9:1429899. doi: 10.3389/feduc.2024.1429899

COPYRIGHT

© 2024 Lange-Kuettner. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Are school grades correlated with competencies in secondary school pupils with special needs?

Christiane Lange-Kuettner*

INSIDE Project, Department. 1, Leibniz Institute for Educational Trajectories (LIfBi), Bamberg, Germany

Introduction: The current correlative study investigates whether and to what extent school grades are related to competencies in adolescents with and without special educational needs and disabilities (SEND) in inclusive secondary education.

Methods: A sample of $N = 2,998$ adolescents with a mean age of 12 years were longitudinally assessed in German language and Mathematics in a nationwide project on inclusive schooling in Germany in 2018/19 (T1) and 2019/20 (T2) in secondary school. The hypothesis was that competencies and school grades should be significantly correlated in both adolescents with and without SEND, showing the reliability of school grades for either group.

Results and discussion: Statistical analyses showed (1) all adolescents improved their competencies independently of their SEND status while school grades were moderately stable over time, (2) more variability of competencies and school grades emerged only at the tail ends of the scales of fail and best scores, (3) correlations between competencies and grades were consistently higher in mathematics than in German language for both pupils with and without SEND indicating a more objective and reliable measure.

KEYWORDS

assessment, competencies, disability, inclusive education, special educational needs and disabilities (SEND), mathematics

Introduction

The current investigation examines the correlation between school grades and competencies in secondary school pupils with and without special educational needs and disabilities (SEND). School grades are assessments by teachers and are multi-dimensional constructs that include teachers' evaluations of effort, engagement and participation that are highly predictive of school success (Bowers, 2011; Brookhart et al., 2016). In this way, they encompass more than ability and competence (Lettau, 2021). Children already comprehend the concept of competence as an assessment of ability (Stipek and Iver, 1989), but the underlying causes they imagine are shifting from effort, social reinforcement, and mastery to objective and normative information. In the current study, particular competencies in math and language (Korhonen et al., 2012) were measured with the aim to test whether the grades and competencies in these subjects correlate in mainstream pupils in secondary schools and also in those with special educational needs and disabilities (SEND).

It is important to look at the school grades and competencies in these two groups because in 2006, the United Nations have released the UN convention on the rights of persons with disabilities which includes the right to attend mainstream schools thereby

participating in general education 'Respect for the evolving capacities of children with disabilities and respect for the right of children with disabilities to preserve their identities' (United Nations, 2008). While inclusive education is now a legal requirement in European states, educational systems face several challenges when implementing such changes at the national level (Lang et al., 2011) as well as at the federal level in Germany (Lange, 2017). Progress in Germany has been slow because in some federal states, school administrations are keeping special schools in place in parallel to inclusive schools (Lange, 2017) as indicated by the fact that only 524 of 3,330 special schools were shut down in 12 years between 2008/9 and 2020/1 (Klemm, 2022). There are several strategies for the implementation, though, for instance, gradually reducing selective school readiness tests in favor of assessment of the need for educational learning support (Kastner-Koller and Deimann, 2018) or to change the function of existing special schools into assessment and support centers (Klemm, 2022). Also, changes in teaching staff were necessary, with the hiring of school assistants as classroom support, and the transfer of special education teachers to regular schools as cooperation partners with mainstream teachers (Grosche and Volpe, 2013; Lange, 2017).

Grading in inclusive secondary schools

A fundamental aim of standardized educational assessment is to achieve reliable identification of students through assessment of skills and abilities, especially numeracy and literacy (Howard et al., 2017). To this end, teachers need to create a cohesive classroom atmosphere where pupils like each other and do not exclude those whom they perceive as weaker or less clever (Gamboa et al., 2021). To achieve this, teachers learn to reach out both to other colleagues and to families and communities (Miller et al., 2022). At the same time, the goal is to uphold academic standards or to even accelerate academic development when pupils with SEND need to be accommodated (Dare and Nowicki, 2023). One strategy is to develop personal development and achievement plans for pupils with SEND (Lange, 2017) and to use differentiated instruction (Nusser and Gehrer, 2020), for instance by marking of pupils' work using different scales. However, to-date research on adjusted marking scales is nearly non-existent, presumably because it is more common to provide additional support to reach common standards (Carter et al., 2022). The school marks of pupils with SEND are reported to be lower than those of the mainstream children, beginning already in primary school and presumably continuing thereafter, which shows no indication of an adjusted marking scale that would make allowances (Parsons and Platt, 2017).

A review showed that to evaluate academic abilities in populations with diverse special educational needs is challenging because of the differences in methodology used in assessments (Evans et al., 2017). Based on Festinger (1954), Bosch (2023) differentiates between the factors task, authority and evaluation structure as being involved in assessment. In the evaluation structure, there are two functions: self-improvement and self-enhancement. Bosch explains that downward comparisons, frequent on-the-spot comparisons, and a lower perceived importance of academic achievements are common among low achievers. In order to compensate for a tendency to assign lower marks to pupils with special educational needs when marking them,

teachers must modify expectations and criteria in their performance-related marking scales.

The current study

The current study is assessing whether the marks of pupils with SEND reflect their competencies to the same degree as in mainstream pupils. The study was conducted in Germany where a SEND diagnosis can be based on one or more of the criteria learning difficulties, emotional and social developmental problems, language problems, physical and motor developmental delays, mental development, problems with hearing and vision, and autism (Gebhardt et al., 2015).

Both the competencies and grades of pupils could be compared starting in the second year of secondary school. It could be assumed that a degree of leniency and consideration of effort (Jung, 2008) would produce somewhat lower correlations between competencies and grades in pupils with SEND rather than being non-significant. This would be the case because grades must not be arbitrary as also schools with inclusive schooling are assessed and often part of large-scale evaluations (Welsh and D'Agostino, 2008).

The large-scale study of grades and assessment of competencies was carried out in a nationwide research project to evaluate the results of inclusive education on a longitudinal basis at the secondary school level (Schmitt et al., 2023; Schmitt et al., 2020) in Germany. The current longitudinal correlative study analyzes the development of competencies and school grades in German and Mathematics only in secondary schools with inclusive schooling in Germany (INSIDE project) in mainstream pupils and those with SEND (Schmitt et al., 2020). The INSIDE project was launched in 2016 when the current state of educational inclusion in Germany needed to be explored to ascertain whether Germany had achieved the aim of the UN Convention on the Rights of Persons with Disabilities to include pupils with special educational needs into the mainstream school system. The first phase of the project focused on investigating the conditions that make inclusive learning successful, which are crucial for developing political strategies and reform measures. Until 2021, main questions of the INSIDE project were for instance, first, what are classroom processes that contribute to the successful individual development of students with special educational needs, and second, what effects does inclusion have on fellow students without special needs? In a second phase of the INSIDE project, from 2021 until 2025, as the school cohort was nearing completion of secondary school, further research questions were how pupils who were taught inclusively would envisage their further life trajectory. In the current study, it is investigated whether in pupils with and without SEND, competencies as measured by educational tests in German language and mathematics are correlated with the grades that teachers give them in these subjects.

Brookhart et al. (2016) describe in a very thorough review that grading developed from oral evaluation into an increasingly standardized assessment, so much so that some in education were equating grades with competencies and were asking for reliability to be as high as in psychometric assessments. US studies found a correlation between standardized tests and school grades to be around .50, as grades are considered 'academic enablers'. School grades were found to be more reliable predictors of school success than psychometric tests that measure competencies. Bowers (2011) suggested that effort, engagement and participation would contribute

to a school grade in addition to academic ability. Other academic enablers were punctuality and quality of homework, oral participation during school hours, application of learning strategies and the ability of pupils to work together in groups. Hence, it is assumed here that competencies and school grades represent different but overlapping areas in development and that the current study will deliver new results on the degree of overlap in inclusive schools in Germany.

Previous research suggests that especially school grades in German language are susceptible to factors beyond competence and academic enablers. For instance, teachers with high expectations and a positive attitude toward ethnicities also have pupils with better marks while this was not the case in mathematics (Peterson et al., 2016). Moreover, we find the old acquaintance of gender for language grades with a large effect size of -0.94 predicting better grades for girls and to a lesser degree the mother's university education (-0.19) as well as a migration background (0.26) and being on an academic school track (-0.33) (Bayer et al., 2021). Most relevant for the current study, in this study the variable special educational needs showed an effect size of 0.71 , independently of the number of students with SEND in the school and the teacher variables. Bayer et al. (2021) state that this effect could not be interpreted meaningfully because the assessment process varies.

Thus, the first hypothesis was that also in the current study, pupils with SEND would have lower school grades and competencies. The second hypothesis was that nevertheless, their competencies scores also improve in this longitudinal data set. The third hypothesis was that their competencies systematically correlate with school grades as in mainstream pupils, although perhaps not to the same extent. Such a result would imply that while the assessment process of pupils with SEND may vary, the school grades would still be systematically related to related to their abilities in the two school subjects, with higher correlations in mathematics than in German language, albeit to a lesser degree than in the pupils without SEND.

Methods

Participants

The original sample size was $N=3,385$, with 1,708 boys (50.5%) and 1,677 girls (49.5%). At the beginning of the longitudinal data collection in the INSIDE project in 2018/9, adolescents were between 10 and 15 years, with most pupils at age 12 (65.1%). For 10.7% or 359 teenagers, German was not the first language. The number of participating schools nationwide delivering data on the school subjects Mathematics and German language decreased from 214 schools in 2018/19 ($N=3,385$ pupils) to 45.6% of the participants in 102 schools ($N=1,544$ pupils) in 2019/2020 during the first wave of COVID-19 when schools were closed because vaccinations were not yet available.

The data used in the current study were collected longitudinally for the school years 2018/19 (T1) and 2019/20 (T2) from $N=2,999$ adolescents. There were 1,519 boys (50.7%) and 1,479 girls (49.3%). Pupils were given T2 grades before the COVID-19 epidemic in January/February 2020, but data on grades were collected from May to July 2020 when schools were closed and pupils had to carry out home schooling. The data set of one adolescent was excluded due to missing values for the competencies variable at both points of measurement. The mean age of the remaining sample ($N=2,998$) at T1 was $M=12;7$ (years; months), with $SD=7$ months. For three data

sets, no information about the special educational needs status was available.

When comparing the three groups, that is pupils without SEND, pupils with SEND diagnosed and pupils with SEND assessed by the school, it turned out that there were 1,328 boys (49.3%) and 1,368 girls (50.7%) pupils without SEND, 156 boys (61.9%) and 96 girls (38.1%) pupils with a SEND certificate and 33 boys (70.2%) and 14 girls (29.8%) with a school assessment of SEND. Chi-square analysis showed that there were significantly more boys in the two SEND groups, $\chi^2(2, 2,995)=22.06, p<0.001$.

With regards to the number of languages spoken, data was not available for some pupils. ANOVA showed that there was no significant difference, $F(2, 2,565)=.01, p=.999$ as the majority of adolescents spoke only one language (without SEND 87.4% of 2,565 pupils, SEND certificate 87.2% of 228 pupils, SEND school 87.2% of 39 pupils) and also the number of pupils speaking a second language was similar (without SEND 12.4%, SEND certificate 12.7%, SEND school 12.8%).

Likewise with regards to the educational background of the pupils, data was not available for some pupils. The length of education in years of the father was not significantly different in the three groups, $F(2, 1463)=.03, p=.967$, as the years in education was very similar (without SEND $M=11$ years 8 months of 1,350 pupils, SEND certificate $M=11$ years 8 months of 98 pupils, and SEND school $M=11$ years 4 months of 15 pupils). The length of education in years of the mother was also not significantly different in the three groups, $F(2, 1545)=1.07, p=.345$, as their years in education only varied slightly (without SEND $M=11$ years 11 months of 1,428 pupils, SEND certificate $M=10$ years 11 months of 101 pupils, and SEND school $M=13$ years 0 months of 16 pupils).

The Bundesland where pupils lived are known but it is prohibited by the German ministry for research and education who financed the data collection to analyze differences between the federal states in Germany. For the interested reader, there is a report available that summarizes the different policies and strategies of implementing inclusive education in each Bundesland (Lange, 2017). Moreover, because there were no significant differences between the parental educational and migration background, these variables were not included in the statistical analyses. Gender could have been included in the statistical analyses, but first, sex differences were not part of the hypothesis, and second, the sample size of the pupils with SEND based on a school assessment was too small to be split up.

Measures

Competencies

The assessment of competencies was not part of regular school assessments. Instead, for the nationwide assessment, extra-curricular specially developed tests were given across parts of the country (Bundesländer). German language competencies were determined through non-fiction texts, advertising texts, literary texts and instructional text comprehension (Berendes et al., 2013). Mathematics competencies were assessed with established mathematics tests (Bos et al., 2009). The language test lasted about 30 min and the math test about 45 min. The variables of the competency measures are based on Item Response Theory and are Weighted Maximum Likelihood Estimates (Warm, 1989). Scores were normally distributed.

School grades

School grades were entered in questionnaires by the schools. Grades that used point systems were recoded into the common and well-known grading scale from 1 to 6 using publicly available conversion tables. A grade of 1 is very good, 2 is good, 3 is satisfactory, 4 is sufficient, 5 is fail (with compensation opportunity by better grades in other subjects when considering progression) and 6 is fail (without compensation opportunity). Half-grades and other further grade qualifications through an asterisk or a plus/minus were not taken into account. For analyses purposes, no difference was made between grades of pupils with and those without SEND as they were on a common scale of 1–6 although evaluation criteria may have differed (Bosch, 2023).

Procedure

The data collection was approved by the education resp. culture ministries of the Bundesländer. All schools, teachers, and parents on behalf of their pupils had consented to the voluntary participation in the project. They received a data protection sheet detailing their rights to refuse participation without disadvantage. Data collection was carried out in all Bundesländer apart from Berlin and Brandenburg because of their different school pathways. The data collection was outsourced. Competencies were tested in the classroom by experienced experimenters who were trained by their employer, the International Association for the Evaluation of Educational Achievement (IEA) (Stichting I.E.A. Secretariaat Nederland), Hamburg branch. School coordinators filled in questionnaires detailing, for instance, pupils' school grades, SEND certificate, types of support etc. At the second point of measurement (T2), on-site assessment of competencies was delayed from spring to autumn as schools could be closed because of COVID-19. Yet school grades were entered from May to July as planned but with the questionnaires of 541 pupils (20.85% of *N*) sent by mail to those schools that were closed because of infected pupils or staff.

Study design

This study analyses data from the project 'Inclusion in and after lower secondary tier in Germany' that was supported by the German Federal Ministry of Education and Research (BMBF) under grants IN1503A, IN1503B, IN1503C, and IN1503D to the Leibniz Institute for Educational Trajectories (LIfBi) in cooperation with a nationwide network. A public user data file will be published on project completion in 2025. The author reports that there are no competing interests to declare.

Ethical considerations

The participation in this study was completely voluntary. All consent forms were vetted by the internal data protection department in the Leibniz Institute for Educational Trajectories before they were submitted to the culture ministries of the Bundesländer. Requested

modifications were made accordingly. Consent by school, parents or adolescents could be withdrawn at any time via email to the data collection agency.

Data analyses

Partially missing data was imputed. The imputation method for missing data was automatically selected by SPSS 28.0. Missing values for the school grade variables were supplemented in 100 imputations (logistic regression). Missing values in competencies in German language and Mathematics were imputed 100 times with predictive mean matching (PMM). The three hypotheses were tested with *t*-tests and correlations. Imputed data sets were pooled and analyzed in SPSS 28.0 using the split-file procedure. Power analysis showed that a two-tailed *t*-test requires 100 participants according to Faul et al. (2007). There are two groups of adolescents with special needs and disabilities (SEND), those who have an officially registered SEND certificate (*n* = 252) and those who receive SEND support because the school they attended estimated that this would be beneficial for the pupil (*n* = 47). According to the power analysis, the two groups of pupils with SEND, either diagnosed or assessed by the school, should have been aggregated into one group, but because it was unclear whether these groups were different, the split into two groups was maintained.

Competencies and school grades were longitudinally assessed in grade 6 (T1) and grade 7 (T2) and compared with pairwise *t*-tests (two-tailed). Because of the number of correlations, only *p*-values < 0.001 are indicated in the tables. Effect sizes were not available for pooled data, but the standard error and the confidence interval are tabulated. These longitudinal correlations of competencies and school grades are reported before correlations between competencies and grades at T1 and 2, respectively. Visualizations were produced with JMP (SAS) based on the imported SPSS data spreadsheet. Curves were smoothed (Spline method), not trimmed, and show the means pooled across imputed data sets with the Confidence of Fit as shaded area along the curve.

Results

Longitudinal development

This section tests the first and second hypotheses that pupils with SEND would have lower school grades and competencies but nevertheless, their competencies scores would also improve in this longitudinal data set.

Table 1 shows how competencies and school grades developed in adolescents with and without SEND. Competencies of all adolescents had significantly increased in the following school year, albeit those of adolescents with SEND from a lower baseline. In contrast, the grades stayed the same. The school grades at T2 were always somewhat lower than at T1 but this decrease was not significant. Also correlations between T1 and T2 were significant for competencies, but not for school grades.

Progression trajectories of the three groups between T1 and T2 was further explored with visualizations, with the outcome variable

TABLE 1 Longitudinal group means of the three SEND groups at T1 and T2 (Pearson pairwise *t*-tests, two-tailed).

Competencies							
SEND	<i>n</i>	<i>M</i> T1	<i>M</i> T2	<i>df</i>	<i>t</i>	<i>CI</i>	<i>r</i>
German language							
No SEND	2,696	0.04 (0.02)	0.52 (0.03)	569	-18.92***	-0.53/-0.43	0.67***
SEND CERT	252	-0.73 (0.09)	-0.15 (0.08)	1,132	-7.82***	-0.73/-0.43	0.70***
SEND School	47	-0.81 (0.18)	-0.11 (0.17)	656	-3.72***	-1.07/-0.33	0.60***
Mathematics							
No SEND	2,696	0.07 (0.02)	0.69 (0.03)	435	-25.84***	-0.67/-0.57	0.71***
SEND CERT	252	-0.78 (0.08)	0.00 (0.08)	1,050	-10.62***	-0.93/-0.64	0.70***
SEND School	47	-0.63 (0.18)	0.12 (0.17)	801	-4.47***	-1.09/-0.42	0.66***
School grades							
German language							
No SEND	2,696	2.9 (0.02)	3.2 (0.41)	100	-0.76	-1.13/0.50	0.40
SEND CERT	252	3.3 (0.06)	3.4 (0.40)	107	-0.35	-0.93/0.65	0.34
SEND School	47	3.7 (0.16)	3.7 (0.52)	144	-0.04	-1.06/1.02	0.31
Mathematics							
No SEND	2,696	2.9 (0.02)	3.2 (0.40)	100	-0.76	-1.11/0.49	0.39
SEND CERT	252	3.3 (0.07)	3.5 (0.36)	111	-0.59	0.50/-0.59	0.32
SEND School	47	3.5 (0.18)	3.7 (0.48)	161	-0.36	-1.21/0.83	0.22

SEND CERT = SEND certificate; SEND School = SEND school assessment, *** = $p \leq 0.001$, SE in brackets, CI, Confidence interval lower/upper boundaries.

T2 always on the *x*-axis. Figure 1 visualizes the correlations between T1 and T2 for Mathematics (Figure 1A) and for German language (Figure 1B). The diagonal axes visualize a perfect correlation of 1 between T1 and T2, that is, competencies would not have changed. If a curve would run above the diagonal, competencies would have regressed during COVID-19, for instance a value of 1 slightly above the diagonal at T1 would have changed into a score of 0 at T2. However, curves below the diagonal show that competencies of the group have improved in the following year.

Figure 1 shows that only those adolescents who showed very low competence in either subject were not doing well in the following year at secondary school. Pupils of all three groups could be found at this tail end of the curve. Otherwise, all groups' scores were located below but close to the diagonal which showed that they had improved somewhat. The scores of adolescents with SEND did not considerably differ compared to mainstream teenagers. Of note are also improvements at the upper end of the competency scale, below and further away from the diagonal, which indicated especially large improvements at T2. In comparison, in German language competence development, there were less pronounced changes at the tail ends of the scale.

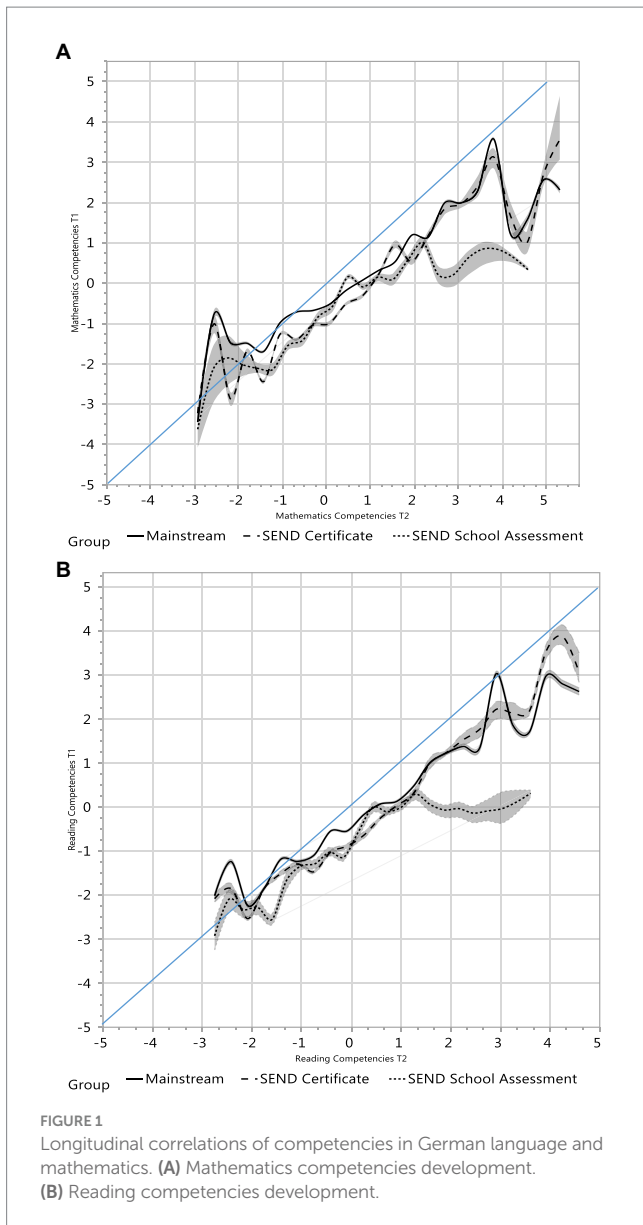
Visualizations of plotted school grades are shown in Figure 2, again with the outcome variable T2 on the *x*-axis. School grades did not change significantly, but they were also not highly correlated (see Table 1). Hence, for school grades a different picture emerged than for competencies. In Figure 2, the scales of the *y*-axes are reversed because 6 is the lowest mark, while 1 is the best mark. Group differences were particularly pronounced at either end, for fail grades and for best grades. This was especially the case for the German language grades.

At the tail end above the diagonal, there are those pupils who had earned a complete fail mark at T2. At the tail end below the diagonal, there are those pupils who had earned the best mark at T2. The spread between groups thus shows that change in a negative or positive way was especially pronounced at the tail ends. Thus, we find that in both Mathematics and German language, mainstream adolescents could still improve their grades from an already good grade, while pupils with SEND earned a best mark even from a lower baseline. There were also mainstream pupils with average grades who failed at T2, while those pupils with SEND who completely failed had already had lower than average marks at T1. In contrast, for the mid-level pass grades that stayed stable (along the diagonal), grade differences between the three groups were reduced. Thus, the low and not-significant correlations between T1 and T2 are explained mainly by changes in the tail ends of the grading scale.

School grades and competencies

This section investigates the third hypothesis that states that competencies systematically correlate with school grades not only in mainstream pupils, but also in pupils with SEND although perhaps not to the same extent. The expectation was that in adolescents with SEND, the correlations between competencies and grades should be lower but still significant and follow the same pattern of results, with correlations between competencies and grades being higher in mathematics. Correlations in Table 2 are listed separately for the three groups.

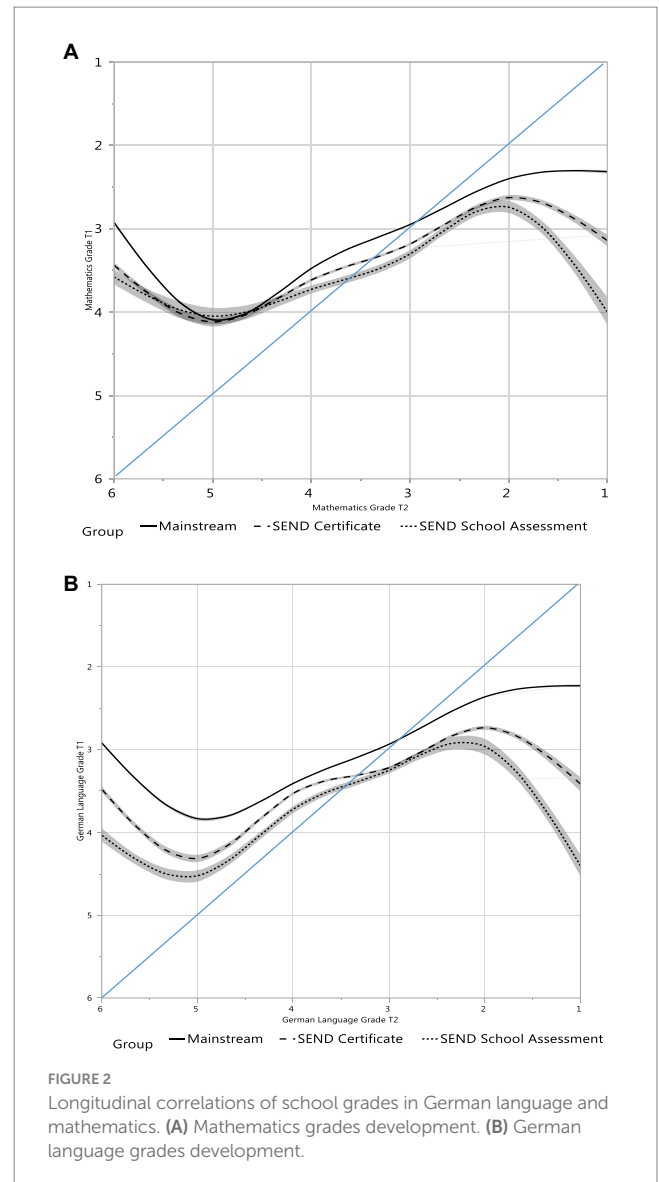
Correlations between grades and competencies are reported for each point of measurement, but also between T1 and T2, because a grade at T1 may also indicate potential at T2. While grades at T1 are



significantly correlated with competencies at T1 and T2, correlations of grades at T2 with competencies at T2 were not reliable for any of the three groups. Correlations of competencies and grades were higher in mathematics than for German language in both mainstream pupils and those with a SEND certificate.

In order to explore why a change in the strength of the correlations between competencies and grades occurred at T2, the correlations at T1 and T2 are visualized in Figure 3. The stronger the slope of the curves, the higher the correlation between competencies and school grades; the more horizontal the curve, the lower the correlation.

In Figures 3A and 3B for T1, the curves show a clear diagonal for mainstream pupils and those with SEND for mathematics. In German language, the curves for adolescents with SEND have a different shape than a diagonal, as again at the tail ends of both the fail and the best grades, differences between the three groups were larger than in the mid-range. Figures 3C and 3D show that at T2, marking appears to be arbitrary and unrelated to their actual competence.



Discussion

The current longitudinal study investigated a first hypothesis that pupils with SEND may have lower grades and competencies but – in a second – hypothesis whether both mainstream and pupils with SEND would develop their competencies and improve their grades in secondary schools with inclusive education. Competencies and school grades are considered to be important areas in development that are different but interrelated (Lettau, 2021). We did find that pupils with SEND have initially lower competencies and school marks in secondary school. Nevertheless, like in mainstream pupils, their competencies improved, while grades were stable. The third hypothesis was that in pupils with SEND, grades should be correlated with competencies like in mainstream pupils although not as strongly. It was interesting to note that the correlations between competencies and school grades were consistently higher in mathematics than in German language for almost all pupils. It may be the case that mathematics’ teachers are more objective and less biased against certain pupils than language teachers as beliefs and opinions may matter less (Bittmann and Mantwill, 2020; Peterson et al., 2016). In

TABLE 2 Domain-specific correlations between competencies and school grades.

School grade	Competencies T1	Competencies T2
No SEND (<i>n</i> = 2,296)		
German language		
German T1	−0.305***	−0.297***
German T2	−0.174	−0.168
Mathematics		
Math T1	−0.392***	−0.440***
Math T2	−0.271	−0.264
SEND certificate (<i>n</i> = 252)		
German language		
German T1	−0.185	−0.186
German T2	−0.143	−0.158
Mathematics		
Math T1	−0.288***	−0.288***
Math T2	−0.230	−0.181
SEND School Assessment (<i>n</i> = 47)		
German language		
German T1	−0.348	−0.307
German T2	−0.080	−0.142
Mathematics		
Math T1	−0.385	−0.386
Math T2	−0.180	−0.153

*** = $p < 0.001$.

confirmation of the third hypothesis, this was true for both pupils with and without SEND. This is even more remarkable given that mathematics was harder to teach during the COVID19 pandemic (Lange-Küttner, 2024). This result confirms the third hypothesis that school grades and competencies in pupils with SEND would be showing the same pattern of correlations albeit to a lesser degree.

An unexpected finding emerged from the visualizations that showed that fail grades and best grades at either end of the scales often showed the most variance between the three groups, while the mid-scale grades and competencies showed less variance. As such, this is a finding that is not in line with the longstanding complaint that teachers' grading is too harsh (Rugg, 1918). Rather than explain this with positive or negative mood which when experimentally induced could influence marking (Brackett et al., 2013), it is suggested here that it indicates that teachers were especially determined to follow their belief that a pupil must be failed or rewarded in order to encourage performance (Lipnevich et al., 2021). Lipnevich et al. (2021) showed that while negative feedback was likely to elicit negative emotions in pupils, there was also an indirect effect that this experience served as a motivational factor to improve performance. Nowadays' tests are or should be usually constructed in such a way that they have questions where the correct and the false answer is clear and for more complex assessments, little room for maneuver and interpretation is given by *a priori* guidance of the expected components of an answer (Vista et al., 2015).

However, it could also be argued from the results of the current study that the competencies' progress analysis also showed that indeed

there was the most change at the tail ends of the scale in either direction, supporting the assumption that at the tail ends of the scales are scores of pupils with the most variable performance, for better or for worse.

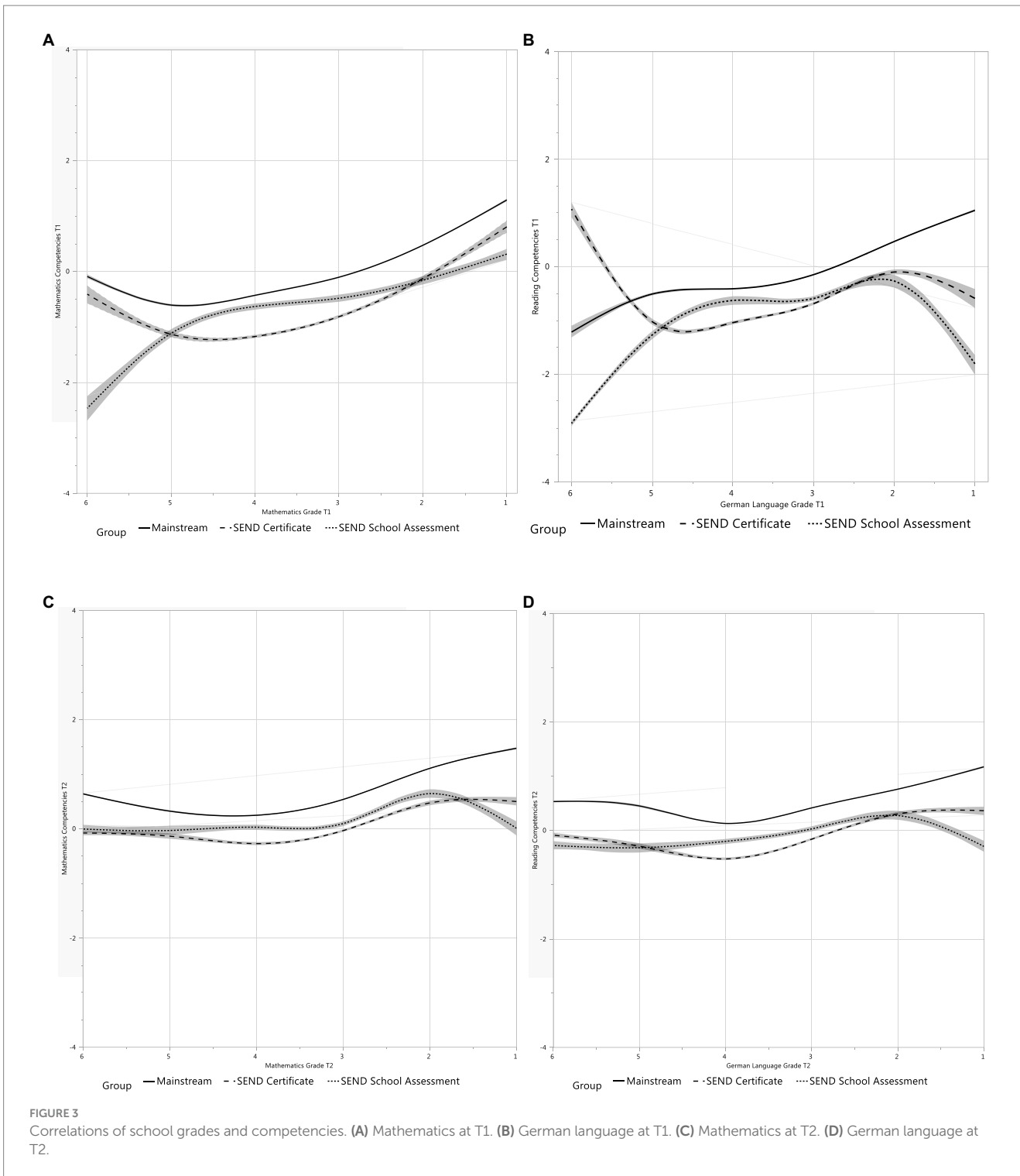
As hypothesized, competencies and school grades were significantly correlated with each other at T1 in both school subjects in mainstream pupils, and as predicted, correlations were higher for mathematics than for German language. For the sample of pupils with a SEND certificate, competencies and school grades were significantly correlated only in mathematics. Moreover, as predicted, correlations in mathematics at -0.29 for pupils with a SEND certificate were significant but somewhat lower than in the mainstream sample with correlations between -0.39 and -0.44 , but both hover around a third of shared variance. Thus, in short, it seems that a more objective marking was available for mathematics than in German language independently of special educational needs.

It must be said, though, that also for the mainstream children, these correlations between competencies and school grades at German schools when compared with those of US schools, where about correlations were hovering around 0.5 (Brookhart et al., 2016), appear to be at the lower end. Also analyses of another large German data sets showed a contingency between competence and school grades was about one third of the variance and thus below 0.5 (Bittmann and Mantwill, 2020) even though additional explanatory variables such as migration and social class background were included (which were more relevant for German language than mathematics). A comprehensive model showed that pupils' motivation was the second most important factor after their actual performance.

However, at T2, curves had flattened, and marking was more arbitrary, especially for pupils with SEND who had considerably improved their competency level but could receive nearly any school grade in either school subject. Thus, the lower and unreliable correlations between competencies and grades in mathematics and German language at T2 were caused by interchangeable marks for pupils with much improved competencies. This could imply that teachers were likely to underestimate the competencies of pupils as they perhaps could not imagine that there can be positive turning points in students' learning in secondary school (Appavoo et al., 2018). Thus, although teachers on average succeeded in lifting most pupils' abilities, this seems less reflected in their grading. It must be noted though that at T2, while the COVID-19 pandemic may not have had an effect on pupils' development of competencies (Nusser et al., 2024), some schools were stressed more than others by the school closures and the transition to digital remote teaching (Lange-Küttner, 2024). Thus, one could suggest that the diminished reliability of school grades across the three groups of pupils at the second point of measurement T2 may have been a consequence of this extraneous factor. Further research would be needed to better understand institutional school stressors. There is data available from interviews with directors of the participating schools in the project, however, their analysis would exceed the limitations of the current study.

Strength and limitations

The current study is part of the large nation-wide INSIDE project on inclusive schooling that investigates how well pupils



with SEND were integrated at secondary school level. This gives the study impressive statistical power as pupils with SEND often consist of very small samples. While a limitation is the larger-than-normal attrition rate during the COVID-19 period, this shortcoming could be ameliorated with multiple imputations. During this period, school grades were less reliably related to the actual competence of pupils. While it does make sense that school grading would suffer during school closures, it also needs to

be considered that teachers have had less observation time of their pupils in the classroom and thus they would have had to make estimates of how much work pupils would have carried out at home. Another strength of the study is that the longitudinal design allowed to prove that the competencies of both pupils with and without SEND improved, even during the COVID-19 pandemic. The fact that both competencies and school grades were longitudinally assessed was especially beneficial as it could

be demonstrated that while grades did not indicate progress, competencies did improve.

Conclusion

The confirmation of the first and second hypotheses that stated that pupils with SEND show lower performance levels but also progression, like the pupils without SEND do, is an important motivation factor for teachers and school administrators to keep a positive outlook on inclusive education. Theoretically, the empirical test could have also revealed a much more negative picture, with mainstream pupils' competence suffering because of reduced teacher attention and pupils with SEND suffering even more because of reduced resources and knowledge about their particular condition. The current study showed that this was not the case and that academic development in inclusive schools was robust even during the pandemic – although school routines such as grading were more disrupted. The third hypothesis that significant correlations between competencies and grades would occur in both pupils with and without SEND was also confirmed. This allows the conclusion that although allowances when grading work of pupils with SEND would have been made in relation to their competence, the adjustment would have been proportionate like for pupils without SEND. An unexpected result was that variability at the tail ends of the school grading scales showed the most variability. An implication for the teaching practice could be that criteria for marking definite fails and outstanding work would need to be better specified by teachers.

Data availability statement

The datasets presented in this article are not readily available because the public user file is work in progress and will be available from May 2025. Requests to access the datasets should be directed to lena.nusser@lifbi.de.

Ethics statement

The studies involving humans were approved by Kultusministerien der Länder der Bundesrepublik Deutschland. The studies were conducted in accordance with the local legislation and institutional

requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

CL-K: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This manuscript analyses data from the project, inclusion in and after lower secondary tier in Germany (INSIDE). This study was supported by the German Federal Ministry of Education and Research (BMBF) under grants IN1503A, IN1503B, IN1503C, and IN1503D to the Leibniz Institute for Educational Trajectories (LifBi) in cooperation with a nationwide network. A public user data file is work in progress and will be published after project completion in May 2025.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author declared that she is an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Appavoo, P., Armoogum, V., and Soyjaudah, S. (2018). Investigating mathematics learning trajectories: a comparative analysis of grades at two major turning points. *Eurasia J. Math. Sci. Tech. Educ.* 14, 1263–1272. doi: 10.29333/ejmste/82537
- Bayer, M., Zinn, S., and Rüdiger, C. (2021). Grading in secondary schools in Germany—the impact of social origin and gender. *Int. J. Educ. Res. Open* 2:article 100101. doi: 10.1016/j.ijedro.2021.100101
- Berendes, K., Weinert, S., Zimmermann, S., and Artelt, C. (2013). Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS)/die Erfassung von Sprachindikatoren über die Lebensspanne im Nationalen Bildungspanel. *J. Educ. Res. Online* 5, 15–49. doi: 10.25656/01:8423
- Bitmann, F., and Mantwill, O. (2020). Gute Leistung, gute Noten? Eine Untersuchung über den Zusammenhang von Schulnoten und Kompetenzen in der Sekundarstufe [good performance, good grades? An analysis of the relationship between grades and competences in German secondary education]. *Soc. Sci. Res. Network*, 1–25. doi: 10.2139/ssrn.3724172
- Bos, W., Bensen, M., Kummer, N., and Frey, K. (Eds.). (2009). TIMSS 2007: Dokumentation der Erhebungsinstrumente zur Trends in International Mathematics and Science Study [Documentation of the assessment instruments of Trends in International Mathematics and Science Study]. *Münster: Waxmann*.
- Bosch, J. (2023). Soziale Vergleichsprozesse in der Schule [Social comparisons in school]. In M. Moritz Börnert-Ringleb, G. Casale, M. Balt and M. Herzog (Eds.), *Lern- und Verhaltensschwierigkeiten in der Schule: Erscheinungsformen-Entwicklungsmodelle-Implicationen für die Praxis [Learning and behavioural difficulties in school: Manifestations and implications for practice]* (pp. 99–107). Stuttgart: Kohlhammer.
- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educ. Res. Eval.* 17, 141–159. doi: 10.1080/13803611.2011.597112
- Brackett, M. A., Floman, J. L., Ashton-James, C., Cherkasskiy, L., and Salovey, P. (2013). The influence of teacher emotion on grading practices: a preliminary look at the evaluation of student writing. *Teach. Teach.* 19, 634–646. doi: 10.1080/13540602.2013.827453

- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., et al. (2016). A century of grading research: meaning and value in the most common educational measure. *Rev. Educ. Res.* 86, 803–848. doi: 10.3102/0034654316672069
- Carter, M., Clements, M., Morris, T., Stephenson, J., Stevens, R., Waddy, N., et al. (2022). The nature of adjustments and monitoring for students with special educational needs in mainstream schools. *Aust. J. Spec. Incl. Educ.* 46, 1–18. doi: 10.1017/jsi.2021.21
- Dare, L., and Nowicki, E. (2023). ‘The road less travelled’: grade-level acceleration in inclusive schools. *Int. J. Incl. Educ.* 27, 188–204. doi: 10.1080/13603116.2020.1837265
- Evans, R., Brown, R., Rees, G., and Smith, P. (2017). Systematic review of educational interventions for looked-after children and young people: recommendations for intervention development and evaluation. *Br. Educ. Res. J.* 43, 68–94. doi: 10.1002/berj.3252
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat.* 7, 117–140. doi: 10.1177/001872675400700202
- Gamboa, P., Freire, S., Anica, A., João Mogarro, M., Fátima Moreira, M., and Vaz da Silva, F. (2021). Correlates of rejection by the peer group: a study comparing students with and without SEN. *Int. J. Incl. Educ.* 28, 381–401. doi: 10.1080/13603116.2021.1941314
- Gebhardt, M., Sälzer, C., Mang, J., Müller, K., and Prenzel, M. (2015). Performance of students with special educational needs in Germany. *J. Cogn. Educ. Psychol.* 14, 343–356. doi: 10.1891/1945-8959.14.3.343
- Grosche, M., and Volpe, R. J. (2013). Response-to-intervention (RTI) as a model to facilitate inclusion for students with learning and behaviour problems. *Eur. J. Spec. Needs Educ.* 28, 254–269. doi: 10.1080/08856257.2013.768452
- Howard, S. J., Woodcock, S., Ehrlich, J., and Bokosmaty, S. (2017). What are standardized literacy and numeracy tests testing? Evidence of the domain-general contributions to students’ standardized educational test performance. *Br. J. Educ. Psychol.* 87, 108–122. doi: 10.1111/bjep.12138
- Jung, L. A. (2008). ‘The challenges of grading and reporting in special education. An inclusive grading model’ in Practical solutions for serious problems in standards-based grading. ed. T. R. Guskey (Thousand Oaks, CA: Corwin Press (Sage)).
- Kastner-Koller, U., and Deimann, P. (2018). ‘Überblick über klassische Verfahren der Schuleingangsdiagnose [overview of classic school entry diagnostic tests]’ in Schuleingangsdiagnostik [school entry diagnoses]. eds. W. Schneider and M. Hasselhorn (Göttingen: Hogrefe), 19–34.
- Klemm, K. (2022). Inklusion in Deutschlands Schulen: Eine bildungsstatistische Momentaufnahme 2020/21 [inclusion in Germany’s schools: a educational statistical snapshot 2020/21]. Gütersloh: Bertelsmann Stiftung.
- Korhonen, J., Linnanmäki, K., and Aunio, P. (2012). Language and mathematical performance: a comparison of lower secondary school students with different level of mathematical skills. *Scand. J. Educ. Res.* 56, 333–344. doi: 10.1080/00313831.2011.599423
- Lang, R., Kett, M., Groce, N., and Trani, J.-F. (2011). Implementing the United Nations convention on the rights of persons with disabilities: principles, implications, practice and limitations. *Alternatives* 5, 206–220. doi: 10.1016/j.alter.2011.02.004
- Lange, V. (2017). Inklusive Bildung in Deutschland. Ländervergleich [Inclusive Education in Germany. Federal state comparison]. Bonn: Friedrich Ebert Stiftung. Available at: <https://library.fes.de/pdf-files/studienfoerderung/13493.pdf>
- Lange-Küttner, C. (2024). COVID-stressed schools struggled to teach mathematics. *Acta Psychol.* 244:104174. doi: 10.1016/j.actpsy.2024.104174
- Lettau, J. (2021). The impact of children’s academic competencies and school grades on their life satisfaction: what really matters? *Child Indic. Res.* 14, 2171–2195. doi: 10.1007/s12187-021-09830-3
- Lipnevich, A. A., Murano, D., Krannich, M., and Goetz, T. (2021). Should I grade or should I comment: links among feedback, emotions, and performance. *Learn. Individ. Differ.* 89:102020. doi: 10.1016/j.lindif.2021.102020
- Miller, A. L., Wilt, C. L., Allcock, H. C., Kurth, J. A., Morningstar, M. E., and Ruppard, A. L. (2022). Teacher agency for inclusive education: an international scoping review. *Int. J. Incl. Educ.* 26, 1159–1177. doi: 10.1080/13603116.2020.1789766
- Nusser, L., and Gehrler, K. (2020). Addressing heterogeneity in secondary education: who benefits from differentiated instruction in German classes? *Int. J. Incl. Educ.*, 1–18. doi: 10.1080/13603116.2020.1862407
- Nusser, L., Lockl, K., Gnams, T., Wolter, I., and Artelt, C. (2024). Geringere Lernzuwächse durch coronabedingte Einschränkungen im Bildungsbereich? [Smaller learning gains because of Corona pandemic constraints on education?], vol. 10: Bamberg: NEPS Corona & Bildung.
- Parsons, S., and Platt, L. (2017). The early academic progress of children with special educational needs. *Br. Educ. Res. J.* 43, 466–485. doi: 10.1002/berj.3276
- Peterson, E. R., Rubie-Davis, C., Osborne, D., and Sibley, C. (2016). Teachers’ explicit expectations and implicit prejudiced attitudes to educational achievement: relations with student achievement and the ethnic achievement gap. *Learn. Instr.* 42, 123–140. doi: 10.1016/j.learninstruc.2016.01.010
- Rugg, H. O. (1918). Teachers’ marks and the reconstruction of the marking system. *Elem. Sch. J.* 18, 701–719. doi: 10.1086/454643
- Schmitt, M., Gresch, C., and Labsch, A. (2023). Challenges in sampling students with and without special educational needs in inclusive settings for an educational longitudinal study. *Surv. Meth. Insights Field* 10:11. doi: 10.13094/SMIF-2023-00011
- Schmitt, M., Roßbach, H.-G., Gresch, C., Stanat, P., Böhme, K., and Grosche, M. (2020). Projekt: Inklusion in der Sekundarstufe I in Deutschland (INSIDE) [project: inclusion in secondary school education in Germany (INSIDE)]. *Erziehungswissenschaft* 31, 199–202. doi: 10.3224/ezw.v31i1.30
- Stipek, D., and Iver, D. M. (1989). Developmental change in childrens assessment of intellectual competence. *Child Dev.* 60, 521–538. doi: 10.2307/1130719
- United Nations. (2008). The UN convention on the rights of persons with disabilities. Available at: <http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf> (Accessed May 06, 2024).
- Vista, A., Care, E., and Griffin, P. (2015). A new approach towards marking large-scale complex assessments: developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review. *Assess. Writ.* 24, 1–15. doi: 10.1016/j.asw.2014.11.001
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika* 54, 427–450. doi: 10.1007/BF02294627
- Welsh, M. E., and D’Agostino, J. (2008). ‘Fostering consistency between standards-based grades and large-scale assessment results’ in Practical solutions for serious problems in standards-based grading. ed. T. R. Guskey (Thousand Oaks, CA: Corwin Press (Sage)), 75–104.