

Development of Self-Concept in Multi-Grade 3rd and 4th Classes

Meike Munser-Kiefer^{a,*}, Sabine Martschinke^b, Alfred Lindl^c, Andreas Hartinger^d

Received : 31 January 2023
Revised : 11 May 2023
Accepted : 21 May 2023
DOI : 10.26822/iejee.2023.305

^a **Corresponding Author:** Meike Munser-Kiefer, University of Regensburg, Germany. E-mail: meike.munser-kiefer@ur.de ORCID: <https://orcid.org/0000-0003-1876-6840>

^b Sabine Martschinke, University of Regensburg, Germany. E-mail: sabine.martschinke@fau.de ORCID: <https://orcid.org/0000-0003-3492-6912>

^c Alfred Lindl, University of Regensburg, Germany. E-mail: alfred.lindl@ur.de ORCID: <https://orcid.org/0000-0002-6969-8385>

^d Andreas Hartinger, University of Augsburg, Germany. E-mail: andreas.hartinger@phil.uni-augsburg.de ORCID: <https://orcid.org/0000-0003-1246-9263>

Abstract

The assumption that multi-grade learning enhances and sustains positive self-concept is widespread, although neither theory nor empirics have yet allowed for firm conclusions. This paper reports on a representative longitudinal study of multi-grade learning in grades 3 and 4 comparing the development of students' self-concept in reading in multi-grade and single-grade classes, also providing a differentiated analysis of the development at varying performance levels. The results show that self-concept is less stable in multi-grade classes. At the end of grade 3, students' self-concept in multi-grade classes is lower than in single-grade classes, although the average achievement level is higher. This effect is mainly ascribed to low-achieving children; however, all students' self-concept recovers by the end of grade 4. Regarding students with a very low self-concept, they are found to a comparable extent in both multi-grade and single-grade classes, but the number of these students is low and the affected students change. The data indicates that pedagogical support is needed, especially in grade 3, to mitigate the effects of social comparison. Further analysis should include the quality of multi-grade teaching implemented in the classroom and its effects.

Keywords:

Multi-Grade; Multi-Age; Stage Classes; Development; Self-Concept

Introduction

For school learning, self-concept is decisive on several dimensions: It is considered part of the students' identity and core of their personality (Haußer, 1995; Choi & Kyung-Hwa, 2021), therefore its promotion has an intrinsic value for basic educational processes in the domain of personality development (Beutel & Hinz, 2008). Moreover, self-concept is considered to play a significant role in achievement development (Arens & Niepel, 2023; Ehm, Hasselhorn & Schmiedek, 2019; Guay, Marsh & Boivin, 2003) and is regarded as a protective factor in challenging situations and whilst facing adversity (Jaarguizar, Garaigordobil & Bernaras, 2018). This is of particular importance for children in the German education system where after four years of joint primary schooling students are usually assigned to different types of secondary schools according to their achievement levels (Martschinke & Kammermeyer, 2003;



Copyright ©
www.iejee.com
ISSN: 1307-9298

Coelho, Marchante & Jimerson, 2017). In general, self-concept is differentiating during the primary school years (Byrne & Gavin, 1996; Ehm, Hasselhorn & Schmiedek, 2019; Marsh, Byrne & Shavelson, 1988; Marsh, 1992). Numerous studies have shown that during these years students' self-concept tend to be unstable and decline (Martschinke & Kammermeyer, 2006; Helmke, 1998; Sewasew & Schroeders, 2019).

Multi-grade learning is supposed to enhance and maintain a positive self-concept that both promotes learning and builds a self-serving self-esteem (summarizing: Veenman, 1995). These assumptions rely, among other things, on the decoupling of self-concept and social comparison (Martschinke & Kammermeyer, 2006) and on the achievement gap in multi-grade tutor-tutee relationships which is perceived as predictable and thus not threatening (Laging, 2010). However, these roles (tutee in the first year, tutor in the second year within the typical cycle of two-years-combination in German multigrade classes) could potentially lead to a destabilization of the self-concept, which could have a differential effect on students of varying performance groups.

This is the primary focus of this paper: On theory and research on self-concept in general and in multi-grade settings in particular, we report results from a representative longitudinal study on multi-grade learning in grade 3 and 4 comparing the development of self-concept in multi-grade and single-grade classes, providing a differentiated analysis of the development at varying performance levels.

Self-Concept as an Important Element of School Learning in Multi-grade Classes

Self-Concept and School Learning

Self-concept refers to ideas, appraisals, and evaluations that individuals have regarding their talents or abilities (Bandura & Wessels; 1994; Moschner & Dickhäuser, 2018). Cognitive representations that relate to school performance are referred to as academic self-concept, which can further be divided into specific subdomains. Shavelson et al. (1976) developed a model of subject-specific classification. Marsh et al. (1988) modified this model differentiating between verbal and mathematical self-concept. Nowadays, this empirically proven model forms the base for numerous empirical studies (Gaspard et al., 2018).

The development of self-concept is influenced by comparison (social, critical, dimensional, and temporal) and by reference group effects (summarized by Möller & Trautwein, 2020). The role of social comparison is empirically well established and there are two main processes that should impact the effect of social

comparison on students' self-concept development: First, the social environment significantly shapes the development of an individual's self-concept by means of direct and indirect feedback (Felson, 1993). This is empirically well established within the context of primary school. Feedback from individuals whose appraisal is perceived as significant, such as parents (Poloczek et al., 2011) or teachers (Spinath, 2004; Ertl et al., 2022) seems of particular importance. However, performance feedback based on formal assessments commonly recognized as a critical and important comparative information fosters an increasingly realistic (and thus lower) self-concept. According to Helmke (1998) this explains why children tend to turn from "optimists into realists" during the first years of primary school (see also Praetorius et al., 2016). Second, social comparisons can also emanate from individuals themselves. In this process students actively compare themselves to other peers, who are subjectively perceived as recognized or well established within the social context. Within school learning, the grade class constitutes one of the main reference groups. Depending on the aggregated achievement level of a specific school class, students of the same achievement level can thus differ regarding the level of their concept (Big-Fish-Little-Pond-Effekt, Köller, 2004, Huguet et al., 2009).

The internal/external frame of reference model (I/E model; Marsh, 1986) and its successor, the 2I/E model (Wolff et al., 2018), describe the development of the academic self-concept. Empirically well supported, it assumes two complementary processes: On the one hand, the individual compares intraindividually distinct dimensions (e.g., mathematics vs. language) or uses temporal information by comparing their own development with their own previous performance. On the other hand, students' self-concept is shaped by interindividual external references based on criteria or social comparison. However, temporal comparisons yield only small effects, whereas dimensional means show medium, and social comparisons yielded the largest effects (Wolff et al., 2018; Wolff et al., 2019). The causal direction of the relationship between academic achievement and (academic) self-concept is not yet fully clarified. Usually, reciprocal effects are assumed (Marsh & O'Mara, 2009).

There are two approaches to explaining the relationship: skill development and self enhancement approach. The skill development approach explains correlations between achievement and self-concept by arguing that performance affects self-concept. By reflecting and classifying one's own performance, the learner recognizes indications for success or failure and draws inferences of this information to form his or her self-concept. The self-enhancement approach pursues the opposite direction. There, either the motivational and reinforcing effects of a positive self-

concept are emphasized more strongly or its negative effects when the expected ability to solve a task is low due to an also low self-concept (e.g., Möller and Trautwein, 2020; Moschner and Dickhäuser, 2018).

Most findings for primary school show that at the beginning (probably also due to the few school experiences), the self-concept tends to influence academic performance initially, but this turns in the further course (e.g., Praetorius et al., 2016; Renner et al., 2011). However, there are also studies that – partly depending on the statistical procedures used – obtain findings which deviate from this (Ehm et al., 2019), accompanied by the stability of the students' academic self-concept. Here, a high stability (especially in comparison to classmates) is evident (summarized: Möller & Trautwein, 2020), although the academic self-concept generally decreases (Sewasew & Schroeders, 2019). In addition, students perceive their own abilities in different domains in an increasingly differentiated way.

Self-Concept in Multi-grade Settings

While the body of research on self-concept and its development is solid and empirically well-founded, it contains only few studies that specifically focus on self-concept (development) in the context of multi-grade settings. Most studies on multi-grade settings typically focus on achievement revealing a general scientific consensus among researchers on the lack of overall achievement differences between mixed- and single-grade classes (e.g., the meta-analyses of Gutiérrez & Slavin, 1992; Russel, Rowe & Hill, 1998, Ronsley-Pavia, Barton & Pendergast, 2019).

In contrast, effects on students' self-concepts can be expected due to the changing frame of reference and comparison in multi-grade classes, where social reference is shaped mainly by 2 reference groups (besides the teacher): on the one hand, peers of the same age, on the other hand, older or younger children (depending on the year of attendance in the mixed year group). Depending on the direction of comparison, these groups provide a child with varying feedbacks regarding their performance (Feinman, Roberts, Hsieh, Sawyer & Swanson, 1992). These social referencing processes or reference group effects (Huber, Gebhardt & Schwab, 2015) could thus result in different effects, depending on the attended grade or school performance for instance.

Regarding self-concept, the available findings indicate rather favorable effects of multi-grade learning. This assumption holds true for at least older meta-analyses ((cf. Ford, 1977; Miller, 1991). Pavan (1992) and Anderson and Pavan (1993) also report high overall student scores and favorable effects on self-concept; a finding confirmed by Jungae Park (1996, unpublished

dissertation, cited in Carle 2019) in her meta-analysis of 98 studies from the "Nongraded Programme". Veenman (1995) confirms this trend in his meta-analysis focusing on self-concept in nine studies (from 1971 to 1985). However, his revision of this meta-analysis (1996, 1997) indicates several differential effects: for example, the advance in positive expression of the self-concept in multi-grade classes could be traced back to factors such as a positive pre-selection of children and teachers, or specific teaching strategies (e.g., like individualized teaching settings). The negative effects reported in one of the studies, on the other hand, could be due to teachers being overwhelmed or unfavorable conditions (for instance the grouping of several grade levels with only one teacher).

More recent studies show rather inconsistent results, mostly reporting neutral, sometimes positive, and rather rarely negative effects (see Pistoli, 2018 for a summary).

Research Questions and Hypotheses

When comparing multi-grade and single-grade classes, differences in the development of students' self-concept can be expected due to the different frames of reference, but also regarding to differential effects depending on the year of schooling in the multi-grade class (tutee vs. tutor) and on achievement. Given the paucity of findings on multi-grade classes in grades 3 and 4, these are important research desiderata.

1) Does self-concept development differ in multi-grade and single-grade classes?

In the multi-grade setting, the frame of reference (Marsh, 1986; Wolff et al., 2018) regarding possible achievement is wider than in the single-grade setting. The extent to which differences in self-concept can be found between mixed- and single-grade classes could depend on the frame of reference students use for their comparisons (Köller, 2004, Huguet et al., 2009). It can be assumed that the students are not exclusively oriented towards children of the same year of school attendance. Therefore, differential effects can be expected for the first and second years of school attendance: In the first year of attendance of multi-grade settings, self-concept could drop, as children also compare themselves with older classmates. In the second year of attendance, it is likely that self-concept scores will recover and thus improve, as at that point social comparison will also take place with the younger children (Wolff et al., 2018; Wolff et al., 2019).

1) Are there differences in the stability of students' self-concept?

In the multi-grade setting, the reference extended by one school year is expected to weaken the self-concept (Sewasew & Schroeders, 2019). Therefore, we expect self-concept scores to be less stable than in single-grade classes (summarized by Möller & Trautwein, 2020).

2) Are there differential effects for children with different learning backgrounds?

Assuming an extended frame of reference (Marsh, 1986; Wolff et al., 2018) in the mixed-year setting, it can be expected that through the direct comparison with students of the higher school attendance year, especially higher-performing students recognize their own weaknesses, which they did not perceive before. This effect is presumably less pronounced for lower-performing students, as they are accustomed to seeing other children perform better.

For the second year of school attendance, a recovery of self-concept is more likely for students of medium and higher performance in the multi-grade setting. For students with low learning performance, it is less clear in which direction the effect spreads: They could either benefit from being the older and thus more competent learning partner or feel threatened with regard to their self-concept. This might happen if they notice – because of the comparisons (Huber, Gebhardt & Schwab, 2015) – how small the gap to the younger students is or that they might even be overtaken by them.

3) Are there differences in students' development if there is a very low self-concept at the beginning of grade 3?

The effect's unclear direction also affects students with extremely low self-concept scores who are taught in multi-grades classes: More flexible and less self-esteem-threatening references between children could foster a positive development of students' self-concept with additional opportunities, especially in the second year. However, the perceived large gap in performance could also stabilize the extremely low level.

Method

Design and Sample

The study involved 1,644 students from 125 classes (68 of which were multi-grade) at 58 elementary schools taught by 125 teachers (91.7 % female; mean seniority MW = 15.8 years, SD = 11.3). The regions were two large cities in southern Germany and the corresponding rural regions of the administrative districts. Recruitment was based on the sample of multi-grade classes. It

was possible to recruit 90% of the multi-grade classes to participate. Reasons for refusal were heterogeneous and unsystematic, therefore a representative sample can be assumed. Representatives of the responsible school supervision and administration were included in the selection of the single-grade classes in order to obtain schools with a comparable district and teachers with a comparable level of competence. A larger control group was intentionally targeted to be able to deal with possible sampling bias by means of suitable matching procedures.

Investigation tools

The study focuses on "reading" as a central aspect of learning in primary education. In addition, reading is present in all lessons, so one can expect multi-grade learning opportunities, whereas children from different years of school attendance might be taught separately in other school subjects.

Self-concept in Reading

The self-concept was conceptualized as a cognitive representation of one's own abilities. Self-concept of reading was chosen because it is likely to be affected by multi-grade teaching. The corresponding questionnaire scale comprised five items (Martschinke & Kammermeyer, 2003), which were elicited in a child-friendly procedure using the four-item Harter scale (Harter, 1990). Trained test administrators were given a standardized instructional guide in advance and practiced the scale format with the children using child-appropriate examples. They read the questionnaire aloud so as not to confound students' response behavior by reading ability. The children took the self-assessments per item in two steps guided by the test administrators: First, the children were asked to make a directional decision and then an expression decision (e.g., Step 1: "Do you find reading easy or difficult?", Step 2: "Do you find reading very easy or only a little easy?" or "Do you find reading very difficult or only a little difficult?").

The students' self-concept in reading was measured three times: the beginning of grade 3 (t_1), the end of grade 3 (t_2), and the end of grade 4 (t_3). Internal consistencies of the scale were adequate at all three time points (Cronbach's alpha, 5 items, $\alpha_1 = .82$, $\alpha_2 = .86$, and $\alpha_3 = .87$).

Performance in Reading

To measure the performance in reading, the same test from the nationwide VERA 2006 school achievement study was used at the beginning of grade 3 and at the end of grade 4 (continuous non-fiction text, closed and open response format, subscales: lower hierarchy and

higher hierarchy comprehension processes, 13 items, Cronbach's $\alpha_{11} = .73$; $\alpha_{13} = .72$). At the end of grade 3, reading performance was measured using a Bavaria-wide comparison test (Lankes, Rieger & Pook, 2015).

Covariates

All covariates were assessed at the beginning of grade 3 via questionnaires using individual items (e.g., gender, number of books in the household for educational background, parental or family language as the main language/ first language of communication between the respective family members) or corresponding scales. Each of these showed good reliabilities (attitude toward classmates and school, 8 items, $\alpha_{11} = .82$; self-concept in mathematics, 10 items, $\alpha_{12} = .87$) (see Martschinke & Kammermeyer, 2003 for the individual scales). Motivation, with the motivation styles intrinsic, identified, introjected, and external – which were also surveyed in the Self-Regulation-Questionnaire (Ryan & Connell, 1989) – was assessed with the help of a dominance-pair comparison (cf. Hartinger et al., 2004). For this purpose, two items were formulated for each of these four motivational styles, resulting in a total of twelve pairwise comparisons (e.g., In class I cooperate a) because I would be ashamed if I did badly or b) because school is very important). Children then had to choose one of the two options. As a measure of the consistency of such pairwise comparisons, Bortz, Lienert, and Böhnke (2008, pp. 489ff.) suggest calculating a characteristic value based on the inconsistent triads which should be avoided. In this case none of the dominance pairwise comparisons has inconsistent triads, hence the dominance pairwise comparisons can be regarded as reliable.

Data Processing and Analysis

Missing values for one or more variables, the proportion of which was 14.7% or less, were estimated for each time point using the Expectation Maximization-algorithm (Enders, 2010, Madley-Dowd et al., 2019). The missing values of the students were first estimated for each time point individually, then across the different time points, if a minimum of 70% of all variables was available for this purpose.

To statistically control for as many covariates as possible, a 1:1 propensity score matching procedure was then conducted using a nearest neighbor-algorithm (Guo & Fraser, 2015). For this purpose, a logistic regression model with the dichotomous criterion single- or multi-grade was estimated, into which various characteristics such as gender, educational background, parental language, family language, and motivational aspects entered as independent variables (McFadden's Pseudo $R^2 = .01$; cf. for details Munser-Kiefer et al., 2021). As a result, the small

differences from a priori between the control and experimental groups with respect to the covariates could be almost completely compensated (mean absolute standardized difference in the matched sample .01) and each child in a multi-grade class ($n_{mg} = 663$) could be assigned exactly one other child with similar characteristics in the single-grade group.

To be able to compare the differing raw score scales of the reading tests used at the three different time points, they were z-standardized after the matching procedure. The following analyses and tables (e.g., Table 1) refer to these z-scores, which also form the basis for the division into four performance quartiles in reading at the beginning of grade 3 to examine research question 3 (Table 4).

Determining developmental trajectories and differences in self-concept between single- and multi-grade groups overall (cf. Tables 1 and 3) as well as for the four performance-based quartiles (Tables 4 and 5) at the three time points, linear mixed models are estimated. These consider the longitudinal data structure (cf. the intraclass correlations [ICC] in Tables 3 and 5) and have further methodological advantages with regard to analysis requirements, power, or the handling of missing values (cf. for details Hilbert et al., 2019). The predictors group (single-grade vs. multi-grade) and time (beginning or end of grade 3, end of grade 4, respectively) are dummy-coded (0/1), with single-grade and end of grade 3 being the reference categories. This is advantageous because in a single model (i.e., without alpha error accumulation) effects between two the groups as well as between the time points can be estimated directly in pairs. The respective interaction effects (group \times beginning of grade 3 or group \times end of grade 4) are of particular interest, as these express the additional change in self-concept in the multi-grade group (considering the change in the single-grade group).

The analysis requirements (e.g., normal distribution) were checked graphically and by inferential statistics and do not limit the interpretability of the results. All further analyses were performed with the statistical software R (R CoreTeam, 2021) and using the following packages: MatchIt (Ho et al., 2011), ggplot2 (Wickham, 2016), multilevel (Bliese & Bliese, 2016), lme4 (Bates et al., 2014), lmerTest (Kuznetsova et al., 2017), and MuMIn (Barton, 2020).

Results

Overall effects of multi-grade teaching on self-concept in reading

A first overview on the development of the reading self-concept in grades 3 and 4 is given in Figure 1 and Table

1, which also shows the corresponding means of the reading performance for comparison purposes (for an analysis of the performance development in reading Munser-Kiefer et al., 2021). Given the underlying four-point rating scale (0 = strongly disagree, 3 = strongly agree), students in both groups show relatively high self-concept on average at all three time points. The distributions are also comparable in both groups at each time point, as corresponding Levene tests show ($F_{\text{beg3}}(1, 1324) = 2.58, p_{\text{beg3}} = .11; F_{\text{end3}}(1, 1324) = 0.33, p_{\text{end3}} = .57; F_{\text{end4}}(1, 1324) = 1.13, p_{\text{end4}} = .14$).

Looking at the developmental course, the values are on average the lowest at the end of grade 3 (cf. Fig. 1 and Tab. 1#). This effect is mainly due to the multi-grade classes: While multi-grade and single-grade classes show almost the same means in self-concept

at the beginning of grade 3 due to the matching procedure, there is a descriptive difference to the disadvantage of multi-grade classes at the end of grade 3 ($d_{\text{sg-mg}} = -0.11$). However, this reverses in favor of multi-grade classes at the end of grade 4 ($d_{\text{sg-mg}} = 0.10$).

The descriptive evidence that classroom organization has an impact on the development of self-concept can be concretized using mixed linear models with the dependent variable self-concept and dummy-coded predictors, with end of grade 3 and single-grade groups as reference categories (see Table 2). Again, we find that the decline in self-concept is somewhat larger in multi-grade classes during grade 3 and increases (compared to single-grade classes) in grade 4, as suggested by the corresponding significant interaction effect "MG x end of grade 4" in Table 2.

Table 1

Self-concept and performance (z-standardized) in reading at three time points (N = 1326)

| M (SD) | Beginning of grade 3 | | | End of grade 3 | | | End of grade 4 | | |
|--------------|----------------------|-------------|------------------------------|----------------|-------------|------------------------------|----------------|--------------|------------------------------|
| Self-concept | 2.39 (0.51) | | | 2.33 (0.55) | | | 2.40 (0.53) | | |
| Performance | 0.00 (1.00) | | | 0.00 (1.00) | | | 0.00 (1.00) | | |
| M (SD) | SG | MG | $d_{\text{SG-MG}}$ [95 % CI] | SG | MG | $d_{\text{SG-MG}}$ [95 % CI] | SG | MG | $d_{\text{SG-MG}}$ [95 % CI] |
| Self-concept | 2.40 (0.52) | 2.38 (0.50) | -0.04 [-0.14; 0.07] | 2.36 (0.55) | 2.30 (0.56) | -0.11 [-0.21; 0.01] | 2.37 (0.54) | 2.42 (0.52) | 0.10 [-0.01; 0.21] |
| Performance | 0.00 (1.01) | 0.00 (0.99) | 0.00 [-0.10; 0.11] | -0.13 (1.10) | 0.13 (0.87) | 0.26 [0.15; 0.37] | 0.02 (1.01) | -0.02 (0.97) | -0.03 [-0.14; 0.08] |

N sample size; SG single-grade, MG multi-grade; M mean; SD standard deviation; dSG-MG effect size Cohen's d (according to Cohen, 1992: 0.2 small, 0.5 medium, 0.8 large), CI confidence interval

Table 2

Linear mixed regression model for self-concept in reading using the matched total sample (N = 1326) and considering the longitudinal data structure nested by classes and participants (ICC = 63.9%)

| Fixed Effects | b | SE | df | t | p |
|---|------|-----|---------|--------|------|
| Intercept | 2.36 | .02 | 165.40 | 107.68 | <.01 |
| Multi-grade (MG) | -.06 | .03 | 190.69 | -1.91 | .06 |
| Beginning of grade 3 | .04 | .02 | 2648.00 | 2.52 | .01 |
| End of grade 4 | .01 | .02 | 2648.00 | 0.72 | .47 |
| MG x Beg. grade 3 | .04 | .02 | 2648.00 | 1.67 | .10 |
| MG x End grade 4 | .11 | .02 | 2648.00 | 4.55 | <.01 |
| marg. R ² cond. R ² | | .01 | | | .65 |

N sample size; ICC intraclass correlation; b (unstandardized) regression coefficient; SE standard error, df degrees of freedom; t t-value; p probability of committing a type I error; p ≤ .05 significant, (marg./cond.) R² (marginal/conditional) coefficient of determination. All predictors included are dummy-coded (0/1). Since the time point at end of grade 3 is the temporal reference category, the regression weights concerning the beginning of grade 3 are to be inverted in the interpretation logic.

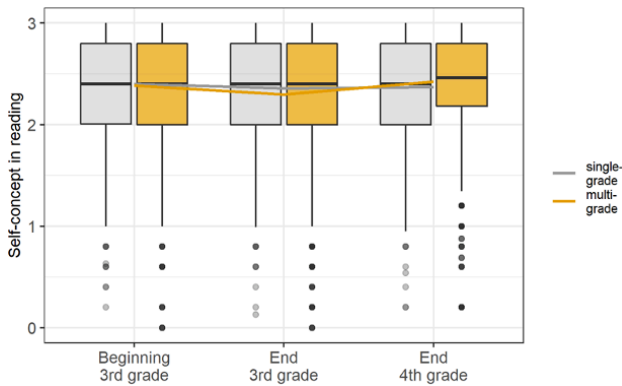
Table 3

Pearson's product-moment correlations between self-concept and performance in reading at three time points (N = 1326)

| | Total | | | Single-grade vs. multi-grade | | | | |
|---------------------------|-----------------|----------------|-----------------|------------------------------|----------------|-----------------|----------------|----------------|
| | Beg. of grade 3 | End of grade 3 | Beg. of grade 3 | Self-concept | | Performance | | |
| | | | | End of grade 3 | End of grade 4 | Beg. of grade 3 | End of grade 3 | End of grade 4 |
| Self-concept Beg. grade 3 | | | | .59 | .58 | .41 | .33 | .33 |
| End grade 3 | .63 | | .67 | | .75 | .41 | .38 | .35 |
| End grade 4 | .57 | .73 | .57 | .71 | | .43 | .37 | .44 |
| Performance Beg. grade 3 | | | .33 | .44 | .49 | | .54 | .51 |
| End grade 3 | .43 | | .28 | .42 | .48 | .36 | | .42 |
| End grade 4 | .55 | .39 | .24 | .33 | .40 | .59 | .39 | |

Above the diagonal multi-grade (grey), below single-grade; for all correlations $r_{ij} p \leq .01$.

Figure 1
Development of self-concept in reading from the beginning of grade 3 to the end of grade 4



Differences in stability of students' self-concept

These observations already provide evidence for stability of the self-concept in reading across the two school years. This is further emphasized by the high correlations between self-concept at the beginning and end of grades 3 and 4 ($r_{\text{beg}3/\text{end}3} = .63$; $r_{\text{end}3/\text{end}4} = .73$; $r_{\text{beg}3/\text{end}4} = .57$). Their contribution even exceeds the corresponding correlation values of reading performance, which are also recorded in Table 3 for comparison purposes. Moreover, a more detailed analysis of Table 3 reveals that the correlation of self-concept (as well as performance) between the beginning and end of grade 3 – but not between other time points – is significantly lower in the multi-grade

Table 4
Self-concept in reading stratified by performance quartiles at three time points

| M (SD) | Beginning of grade 3 | | | End of grade 3 | | | End of grade 4 | | |
|--------------|----------------------|----------------|------------------------------|----------------|----------------|------------------------------|----------------|----------------|------------------------------|
| | SG | MG | $d_{\text{SG-MG}}$ 95 % CI | SG | MG | $d_{\text{SG-MG}}$ 95 % CI | SG | MG | $d_{\text{SG-MG}}$ 95 % CI |
| 1st quartile | 2.11 (0.58) | 2.19 (0.54) | 0.13 [-0.08; 0.34] | 2.05 (0.57) | 1.97 (0.62) | -0.15 [-0.36; 0.06] | 2.07 (0.57) | 2.09 (0.60) | 0.04 [-0.17; 0.25] |
| 2nd quartile | 2.34 (0.52) | 2.30 (0.50) | -0.07 [-0.29; 0.15] | 2.27 (0.54) | 2.24 (0.48) | -0.06 [-0.28; 0.16] | 2.28 (0.48) | 2.32 (0.46) | 0.09 [-0.13; 0.31] |
| 3rd quartile | 2.55 (0.41) | 2.45 (0.43) | -0.24 [-0.45; -0.02] | 2.51 (0.48) | 2.38 (0.50) | -0.26 [-0.48; -0.04] | 2.48 (0.49) | 2.53 (0.44) | 0.10 [-0.11; 0.32] |
| 4th quartile | 2.62 (0.40) | 2.62 (0.41) | -0.02 [-0.23; 0.20] | 2.61 (0.43) | 2.63 (0.37) | 0.06 [-0.16; 0.28] | 2.66 (0.38) | 2.77 (0.26) | 0.35 [0.12; 0.57] |

N sample size; SG single-grade, MG multi-grade; M mean; SD standard deviation; $d_{\text{SG-MG}}$ effect size Cohen's d (according to Cohen, 1992: 0.2 small, 0.5 medium, 0.8 large), CI confidence interval; stratification: 1st quartile group of students with the lowest reading performance, 4th quartile group of students with the highest reading performance

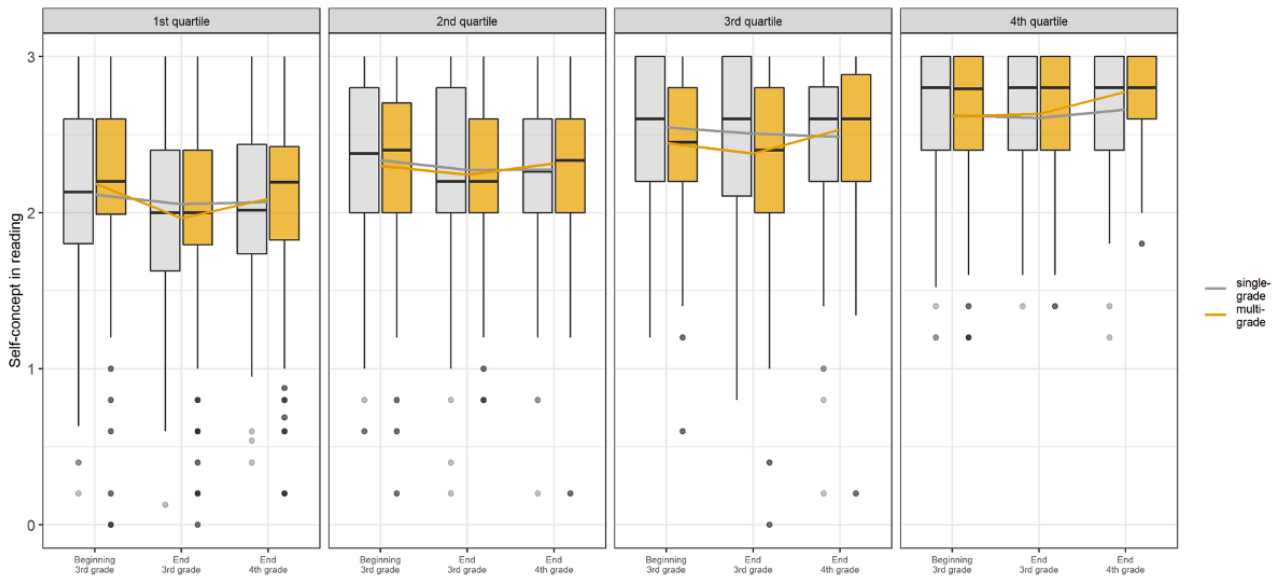
Table 5
Linear mixed regression model for self-concept in reading using the matched sample stratified by performance quartiles and considering the longitudinal data structure nested by classes and participants

| N obs. ICC | 1st quartile | | | | | 2nd quartile | | | | | 3rd quartile | | | | | 4th quartile | | | | |
|---|--------------|-----|--------|-------|------|--------------|-----|--------|-------|------|--------------|-----|--------|-------|------|--------------|-----|--------|-------|------|
| | b | SE | df | t | p | b | SE | df | t | p | b | SE | df | t | p | b | SE | df | t | p |
| Fixed effects | | | | | | | | | | | | | | | | | | | | |
| Intercept | 2.05 | .04 | 561.18 | 46.72 | <.01 | 2.27 | .04 | 602.84 | 58.17 | <.01 | 2.51 | .04 | 635.95 | 67.13 | <.01 | 2.61 | .03 | 590.36 | 90.71 | <.01 |
| Multi-grade (MG) | -.09 | .06 | 561.18 | -1.42 | .15 | -.03 | .06 | 602.84 | -.57 | .57 | -.13 | .05 | 635.95 | -2.53 | .01 | -.02 | .04 | 590.36 | -.59 | .56 |
| Beginning of grade 3 | .06 | .04 | 680.00 | 1.62 | .11 | .06 | .04 | 640.00 | 1.64 | .10 | .04 | .04 | 670.00 | 1.14 | .25 | .01 | .03 | 646.00 | .52 | .60 |
| End of grade 4 | .01 | .04 | 680.00 | .36 | .72 | .00 | .04 | 640.00 | .03 | .98 | -.02 | .04 | 670.00 | -.61 | .55 | .05 | .03 | 646.00 | 1.95 | .05 |
| MG x Beg. grade 3 | .16 | .05 | 680.00 | 3.03 | <.01 | -.01 | .05 | 640.00 | -.10 | .92 | .03 | .05 | 670.00 | .57 | .57 | -.03 | .04 | 646.00 | -.79 | .43 |
| MG x End grade 4 | .11 | .05 | 680.00 | 2.08 | .04 | .07 | .05 | 640.00 | 1.37 | .17 | .18 | .05 | 670.00 | 3.64 | <.01 | .09 | .04 | 646.00 | 2.26 | .02 |
| margin. R ² cond. R ² | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

N sample size; obs. observations; ICC intraclass correlation; b (unstandardized) regression coefficient; SE standard error, df degrees of freedom; t t-value; p probability of committing a type I error; $p \leq .05$ significant, (margin./cond.) R² (marginal/conditional) coefficient of determination. All predictors included are dummy-coded (0/1). Since the time point at end of grade 3 is the temporal reference category, the regression weights concerning the beginning of grade 3 are to be inverted in the interpretation logic; stratification: 1st quartile group of students with the lowest reading performance, 4th quartile group of students with the highest reading performance

Figure 2:

Development of self-concept in reading from the beginning of grade 3 to the end of grade 4 (separated by performance quartiles)



classes than in the single-grade classes ($r_{mg,beg3,end3} = .59$, $r_{sg,beg3,end3} = .67$; $z = 2.42$, $p = .02$). With comparable baseline variance in both groups (cf. Section 5.1), this suggests greater changes in self-concept during grade 3 multi-grade classes and is supported by the greater variance in change in self-concept between the beginning and end of grade 3 compared to the single-grade setting. This difference in variance between the two types of organization was tested and confirmed using Levene's test of difference scores in self-concept at the end and beginning of grade 3 ($F_{sg,mg^*end3-beg3}(1, 1324) = 6.13$, $p = .01$). In contrast, a corresponding analysis of the difference scores between the of grade 4 and grade 3 yields no significant result ($F_{sg-mg,end4-end3}(1, 1324) = 0.28$, $p = .59$). Thus, self-concept proves to be less stable in the multi-grade setting in grade 3 than in the single-grade setting, while a comparable weakening cannot be found in grade 4.

Differential effects in children with different learning prerequisites

Above, we have reported on the comparatively strong decrease of the self-concept in multi-grade classes in the course of grade 3 and, in turn, the increase in grade 4. The question now arises as to whether these effects are consistent or can be attributed to children with specific learning prerequisites. If we look at the descriptive data (Table 4) and the findings of the linear mixed models (Table 5), we see that there are definitely differences: Compared to single-grade classes, the self-concept decreases in the course of grade 3, especially in the group with the lowest reading performance. This effect is significant, as underlined by the interaction term "MG x beginning grade 3" of the associated mixed linear model in Table 5. None of the other performance groups shows a

comparable effect.

At the end of grade 4, the multi-grade classes have in all quartiles on average a noticeably more positive self-concept than the single-grade classes. The higher the performance of the quartile in question, the greater the effects. The changes over the school year are also more positive in the multi-grade classes than in the single-grade classes in all quartiles. They become significant for the first, third and fourth quartile (cf. Table 5 with the respective interaction effects "MG x end of 4th year"). The development is additionally illustrated in Figure 2.

Differences in development of children with very low self-concept at the beginning of grade 3

This question takes on additional relevance because it has just been shown that during grade 3 the self-concept of the children with the lowest learning performance drops particularly sharply.

As a very low self-concept, we define values which are more than three standard deviations below the mean value at the respective time of observation. This category thus includes self-concept means that are less than 0.86 at the beginning of grade 3, less than 0.66 at its end, and less than 0.81 at the end of grade 4. Respectively, this is true for 20, 17, and 18 students. The distribution between the single- and multi-grade group is presented in Table 6.

As can be seen in Table 6, at the beginning of grade 3 and at the end of grade 4, about the same number of students in multi-grade and single-grade classes, respectively, have a very low self-concept. By the end of grade 3, there are almost twice as many children

with very low self-concept in the multi-grade classes. This is consistent with the findings above. At the end of grade 4, the distribution is more balanced again.

Table 6

Distribution of very low self-concept in reading in single- and multi-grade setting

| Beginning of grade 3 | | End of grade 3 | | End of grade 4 | | | | | | | |
|----------------------|----------------|----------------|----------------|----------------|----------------|----|----------------|---|----------------|----|----------------|
| SG | MG | SG | MG | SG | MG | | | | | | |
| N | M (SD) | N | M (SD) | N | M (SD) | | | | | | |
| 11 | 0.60 (0.20) | 9 | 0.42 (0.32) | 6 | 0.42 (0.21) | 11 | 0.35 (0.24) | 8 | 0.54 (0.26) | 10 | 0.45 (0.27) |

SG single-grade, MG multi-grade; N sample size; M mean; SD standard deviation

However, it should also be noted that this group consists of only 7 (4 multi-grade, 3 single-grade) out of the 20 respective children at the beginning and end of grade 3. Only one child (multi-grade) shows a very low self-concept at all three time points.

Summary and Discussion

This paper reports the results of a representative longitudinal study on multi-grade learning in grade 3 and grade 4 with respect to the development of children's self-concept regarding their reading skills. We investigate the question of (differential) differences compared to learning in single-grade classes.

Our data confirm the assumption that at the end of grade 3, self-concept is lower in multi-grade classes than in single-grade classes. At the same time, in these classes the students' self-concept is less stable. Both findings indirectly confirm the assumption that students compare their performance to the entire learning group (in this case extended by one grade).

From this perspective, the findings for grade 4 are consistent: Here, the self-concept of the multi-grade students once more aligns itself to and even tends to be slightly higher than the self-concept of students in single-grade settings. Since the children now belong to the older group, there is a new reference group showing lower learning performance on average. As expected, the more proficient the children are in reading, the more they benefit.

Regarding possible differential effects, our hypotheses were not consistently confirmed: The effect just described concerning grade 3 can mainly be ascribed to the children with low reading performances (at the beginning of the grade 3), and not - as assumed - to students with good and average reading performances. The latter apparently perceive the comparison with older children to be less stressful.

For this reason, it is pedagogically significant that even in the group of children with low reading performances in multi-grade classes, self-concept

increases again significantly in the course of the grade 4. One reason for this could be the fact that especially for those children, switching to the role of the older students (and often tutors) could be especially self-concept enhancing.

Similar effects were found for all performance groups. Again, this can be explained by the extended frame of reference (in this case downwards) within which students compare and classify themselves.

Students with a very low self-concept were found in both settings (multi- and single-grade classes) - the number of children affected, however, is not very high (single-grade: 6-11 and multi-grade: 9-11 of N = 1326). It is striking that there is a roughly consistent number of students with very low self-concept across all times of measurement, however, with one exception as these are not the same students. Overall, the number of students with a very low self-concept indicates that neither form of classroom setting poses an increased risk.

What matches the findings of the other analyses is that multi-grade classes, however, tend to pose a higher risk of weakening students' self-concept - especially in the grade 3, when the comparison with the older children takes place. Our findings thus indirectly confirm that social comparison is highly significant even in multi-grade classes.

In summary, multi-grade settings probably weaken their students' self-concept. However, this is balanced again by the end of grade 4 across all subgroups of reading performance.

This allows for the cautious statement that multi-grade classes, when thought of as a rotation (beginning of grade 3 until end of grade 4), do not have more negative effects on self-concept development than single-grade classes. In fact, multi-grade classes even become an opportunity for a more favorable development - depending on students' performance.

From our findings, indications can be deduced on how to mitigate the less favorable development in the first year of attendance in multi-grade classes as well as on how the favorable effects in the second year of attendance can be supported: For example, students in their first year of attendance could be supported by a corrective based on a relativizing criterion referenced standard as well as by the use of a self-referenced standard with clear, yet still individually set goals.

This could, in addition to the broadened information spectrum induced by social comparison with older students, generate a self-concept protecting effect: first by perceiving a more manageable learning growth (Wigfield & Cambria, 2010) and, second by

perceiving one's own learning growth. Non-public feedback (Dresel et al., 2017) and allocating various tasks to choose from in accordance with one's learning needs could provide additional support and help decouple from social comparison (Martschinke & Kammermeyer, 2003).

Furthermore, the composition of student groups within multi-grade classes is relevant as a reference: Here, a medial difference in achievement could prove to be self-concept supportive. The students in the multi-grade setting would thus have the advantage of developing a more realistic perception of potential and individual goals while at the same time obtaining more favorable motivational effects on their self-concept.

Limitations and Reflections

The study reports effects of multi-grade classes from Germany. Due to the school administration's data protection regulations, it was not feasible to randomize the sampling of classes and students to an experimental design. This was countered by a nearly comprehensive inclusion of multi-year classes from different regions; the single-grade classes of the comparison group were selected according to theoretical criteria. From a methodological point of view, the applied propensity-score matching procedure (see section #) is an appropriate way to control a certain number of covariates (Kuss et al., 2016), even if only actually measured characteristics can be included in the adjustment. This must be considered for all (causal) assumptions and further conclusions.

As the study is restricted to Germany, the teaching (setting) of multi-grade classes is very likely to be more similar within the sample than in the international total population of multi-grade classes. Nevertheless, there are certainly also differences in teaching design and instructional quality, here as well as internationally (Ronksley-Pavio et al., 2019). In our paper, we do not take these differences into account, so we cannot rule out the possibility of rather favorable and rather unfavorable forms of teaching combined into a mean value which may weaken the effects. Since effects on students originate less from overservable structures of learning (e.g., classroom organization as open and guided teaching, multi-grade and mono-grade teaching) than from structures on process level (e.g., cognitive activation, socio-emotional support) (Hahn, 2019), this is a limiting factor. Therefore, further analyses should include the way multi-grade teaching is implemented in the classroom to be able to describe the effects of grade-mixing more comprehensively. In this way, (un)favorable forms of multi-grade teaching can be identified and used for pre-service teacher training and teacher professionalization.

Furthermore, the reported findings on self-concept refer to the cognitive representation of one's own abilities in the field of reading. Thus, the verbal self-concept is narrowed down to a subject-specific facet. This could be extended to the entire spectrum of the verbal self-concept respectively supplemented by facets of self-concept concerning maths and science, which would allow an additional analysis of dimensional effects. In this case and in general, an analysis could be conducted with respect to broader differential effects (such as gender, first language, or socio-economic status), which could complement the data of Quail and Smyth (2014). Additionally, future studies and analyses should consider emotional-affective facets of self-concept, which had also been excluded in this study, as an important elements of personality development.

In order not to overwhelm the children, the self-concept was measured with a short scale (5 items), which describes the underlying construct in a valid manner, but certainly not exhaustively. In addition, the degree of differentiation of the four-item Harter scale (Harter, 1990) is limited, which from the beginning leads to ceiling effects, especially for individuals in the higher ranges of performance. Thus, changes or even larger differences in self-concept between mixed- and single-grade classes may be underestimated or not detected at all (cf. Fig. 2, especially 4th quartile). However, the use of the Harter scale has proven to be successful especially at primary school level, as it is a linguistically and methodologically adapted and simplified procedure for self-assessment of abilities suitable for children (Harter, 1990, Martschinke & Kammermeyer, 2006).

Furthermore, it must be taken into account that at the three survey time points different instruments were used to assess reading performance with different difficulty parameters as well as raw and total scores. This was necessary, on the one hand, with regard to the curricular validity of the tests, and on the other hand to avoid floor or ceiling effects. However, due to the resulting z-transformation, statements can only be made in relation to the total sampling.

Moreover, the consideration of quartiles, which were formed according to the reading performances at the first time of measurement, is ultimately based on pragmatical reasons (e.g. sufficient number of cases per quartile), forms of reading instruction (e.g. clear presentability), and content-related reasons (analysis of the development of low, below-average, above-average, and highest initial performances). In addition, other distributions would also be possible (e.g. terciles); this would lead to comparable results.

Further analyses of the data from our longitudinal study of multi-grade learning in grades 3 and 4 will focus, among other things, on the design of multi-

grade teaching (e.g., with respect to reference norm orientation, to amount of multi-grade learning, and to adaptivity) and its effects on performance, motivation, and attitudes. A special focus will continue to be on differential effects (e.g., students' learning requirements or pre-qualifications, students' length of in multi-grade classes, gender). On the one hand, this will provide information on both opportunities and limitations of multi-grade learning, and on the other hand, it also provides information on how to appropriately design and set this form of classroom setting. Additionally, the children's different learning pre-qualifications and learning needs are emphasized - a fact that is of particular importance in multi-grade learning due to the two grades being taught at the same time.

Funding

This work was supported by Staedler Foundation Nuremberg (non-profit organization) and authorized the Bavarian Ministry of Education, Germany.

References

- Anderson, R. H., & Pavan, B. N. (1993). *Nongradedness: Helping it to happen*. Technomic.
- Arens, A. K., & Niepel, C. (2023). Formation of academic self-concept and intrinsic value within and across three domains: Extending the reciprocal internal/external frame of reference model. *British Journal of Educational Psychology*. DOI: 10.1111/bjep.12578
- Bandura, A., & Wessels, S. (1994). *Self-efficacy*. In V. S. Ramachaudran (Ed.). *Encyclopedia of Human Behavior* (Vol. 4, pp. 71-81).
- Barton, K. (2020). MuMIn Multi-model inference. *R package version 1.43.17*. 2020.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Beutel, S. I., & Hinz, R. (2008). *Schulanfang im Wandel: Selbstkonzepte der Kinder als pädagogische Aufgabe* (Vol. 6). LIT Verlag.
- Bliese, P., & Bliese, M. P. (2016). Package 'multilevel'. *R version, 2*.
- Bortz, J., Lienert, G. A., & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik*. Springer.
- Byrne, B. M., & Gavin, D. A. (1996). The Shavelson Model revisited: Testing for the structure of academic self-concept across pre-, early, and late adolescents. *Journal of Educational Psychology, 88*(2), 215.
- Carle, U. (2019): Jahrgangübergreifendes Lernen. Haring, M., Rohlf, C. & Gläser-Zikuda, M. (Eds.). *Handbuch Schulpädagogik* (pp. 473-482). Waxmann.
- Choi, E. J. & Kyung-Hwa, L. (2021). Analysis of Longitudinal Relationship among Elementary and Middle School Students' Multicultural Acceptance, Self-concept, and Community Consciousness using the Latent Growth Model. *International Electronic Journal of Elementary Education, 13*(4). DOI: <https://www.iejee.com/index.php/IEJEE/article/view/1430>.
- Coelho, V. A., Marchante, M., & Jimerson, S. R. (2017). Promoting a positive middle school transition: A randomized-controlled treatment study examining self-concept and self-esteem. *Journal of youth and adolescence, 46*(3), 558-569.
- Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological bulletin, 112*(1), 155-159.
- Dresel, M., Martschinke, S., Kopp, B., Tobisch, A., & Kröner, S. (2017). Zur Bedeutung des Unterrichts für die Koppelung von sozialer Herkunft und Schulleistung: Ergebnisse einer Studie im Grundschulunterricht im Fach Deutsch. *Zeitschrift für Grundschulforschung, 10*(2), 136-149.
- Ehm, J. H., Hasselhorn, M., & Schmiedek, F. (2019). Analyzing the developmental relation of academic self-concept and achievement in elementary school children: Alternative models point to different results. *Developmental Psychology, 55*(11), 2336-2351.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.
- Ertl, S., Kücherer, B., & Hartinger, A. (2022). Lernentwicklungsgespräche und die Entwicklung motivationaler Aspekte des Lernens. *Zeitschrift für Grundschulforschung, 15*(1), 221-236.
- Feinman, S., Roberts, D., Hsieh, K.F., Sawyer, D., Swanson, D. (1992). A Critical Review of Social Referencing in Infancy. In Feinman, S. (Ed.) *Social Referencing and the Social Construction of Reality in Infancy*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-2462-9_2
- Felson, R. B. (1993). The (somewhat) social self: How others affect self-appraisals. *Psychological perspectives on the self, 4*, 1-26.

- Ford, B. E. (1977). Multiage grouping in the elementary school and children's affective development: A review of recent research. *The Elementary School Journal*, 78(2), 149-159.
- Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology*, 52, 1-14.
- Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of educational psychology*, 95(1), 124.
- Gutierrez, R., & Slavin, R. E. (1992). Achievement effects of the nongraded elementary school: A best evidence synthesis. *Review of educational research*, 62(4), 333-376.
- Hahn, E. (2019). *Umgang mit Heterogenität an Gemeinschaftsschulen: eine multimethodische Untersuchung zu Oberflächen- und Tiefenstrukturen des Unterrichts* (Vol. 6). Münster: Waxmann Verlag.
- Harter, S. (1990). Issues in the assessment of the self-concept of children and adolescents. In A. M. La Greca (Ed.). *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 292–325). Allyn & Bacon.
- Hartinger, A., Graumann, O. & Grittner, F. (2004). „Grundschul-Numerus Clausus“ oder Orientierungsstufe? Auswirkungen verschiedener Übertrittsbedingungen auf Motivationsstile und Leistungsängstlichkeit von Grundschulkindern. *Empirische Pädagogik*, 18, 173-193.
- Haußer, K. (1995). *Identitätspsychologie*. Springer-Verlag.
- Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitskonzeptes vom Kindergarten bis zur 6. Klassenstufe. In F. Weinert & A. Helmke (Eds.). *Entwicklung im Kindesalter* (pp. 115-132). Psychologie Verlags Union.
- Hilbert, A., Petroff, D., Herpertz, S., Pietrowsky, R., Tuschen-Caffier, B., Vocks, S., & Schmidt, R. (2019). Meta-analysis of the efficacy of psychological and medical treatments for binge-eating disorder. *Journal of consulting and clinical psychology*, 87(1), 91.
- Ho, D., Imai, K., King, G., Stuart, E., & Whitworth, A. (2018). *Package 'MatchIt'*.
- Huber, C., Gebhardt, M., & Schwab, S. (2015). Lehrkraftfeedback oder Spaß beim Spiel? Eine Experimentalstudie zum Einfluss von Lehrkraftfeedback auf die soziale Akzeptanz bei Grundschulkindern. *Psychologie in Erziehung und Unterricht*, 62(1), 51-64.
- Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., ... & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of personality and social psychology*, 97(1), 156.
- Jaureguizar, J., Garaigordobil, M., & Bernaras, E. (2018). Self-concept, social skills, and resilience as moderators of the relationship between stress and childhood depression. *School mental health*, 10(4), 488-499.
- Jungae Park, K. (1996): *The impact of the nongraded program on students' affective domains and cognitive domains*. Ed. D. Dissertation: ProQuest.
- Kölller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Waxmann.
- Kuss, O., Blettner, M., & Börgermann, J. (2016). Propensity Score: An Alternative Method of Analyzing Treatment Effects: Part 23 of a Series on Evaluation of Scientific Publications. *Deutsches Ärzteblatt International*, 113(35-36), 597.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82, 1-26.
- Laging, R. (2010). Altersheterogenität und Helfen—eine Untersuchung in der Schuleingangsstufe der Reformschule Kassel. In R. Laging (Ed.) *Altersgemischtes Lernen in der Schule*, (pp. 54-71). Schneider Verlag Hohengehren.
- Lankes, E. M., Rieger, E., & Pook, M. (2015). VERA-3 in Bayern. Ein Instrument für die Schul- und Unterrichtsentwicklung. https://www.isb.bayern.de/download/16050/vera_3_in_bayern.pdf.
- Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of clinical epidemiology*, 110, 63-73.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American educational research journal*, 23(1), 129-149.

- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of educational psychology, 84*(1), 35.
- Marsh, H. W. (2009). This MS is the final prepublication (open access) version of the published article: Marsh, HW, & O'Mara, A. (2010). Long-term total negative effects of school-average ability on diverse. *Educational Psychology, 24*, 51-72.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of educational psychology, 80*(3), 366.
- Martschinke, S., & Kammermeyer, G. (2003). Schulleistung und Fähigkeitsselbstbild im Anfangsunterricht - Ergebnisse aus dem KILIA-Projekt. *Empirische Pädagogik, 17*, 486-503.
- Martschinke, S., & Kammermeyer, G. (2006). Selbstkonzept, Lernfreude und Leistungsangst und ihr Zusammenspiel im Anfangsunterricht. In Schründer-Lenzen, A. (Ed.), *Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang*. (pp. 125-129). VS Verlag für Sozialwissenschaften.
- Miller, B. A. (1991). A Review of the Qualitative Research on Multigrade Instruction. *Journal of Research in Rural Education, 7*(2), 3-12.
- Möller, J., & Trautwein, U. (2020). Selbstkonzept. In E. Wild & J. Möller (Eds.). *Pädagogische Psychologie* (pp. 187-209). Springer.
- Moschner, B. & Dickhäuser, O. (2018). Selbstkonzept. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* (pp. 750-756). Beltz.
- Munser-Kiefer, M., Martschinke, S., Lindl, A. & Hartinger, A. (2021). Leistungsentwicklung in jahrgangsgemischten und jahrgangshomogenen dritten und vierten Klassen. *Unterrichtswissenschaft, 49*. doi: 10.5283/epub.50956.
- Pavan, B. N. (1992). The Benefits of Nongraded Schools. *Educational leadership, 50*(2), 22-25.
- Pistioli, E. (2018). The Self-Concept in Specific Areas of Students attending Multigrade and Single Grade Elementary Schools: Comparative Approach. *International Journal of Learning, Teaching and Educational Research, 17*(12), 116-134.
- Poloczek, S., Karst, K., Praetorius, A. K., & Lipowsky, F. (2011). Generalisten oder Spezialisten? Bereichsspezifität und leistungsbezogene Zusammenhänge des schulischen Selbstkonzepts von Schulanfängern. *Zeitschrift für Pädagogische Psychologie, 25*(3), 173-183.
- Praetorius, A. K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie, 48*(1), 14-26.
- Quail, A., & Smyth, E. (2014). Multigrade teaching and age composition of the class: The influence on academic and social outcomes among students. *Teaching and Teacher Education, 43*, 80-90.
- R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152.
- Renner, G., Martschinke, S., Munser-Kiefer, M., & Steinmüller, S. (2011). Diagnose und Förderung des Selbstkonzepts im Anfangsunterricht. In F. Hellmich (Ed.) *Selbstkonzepte im Grundschulalter: Modelle, empirische Ergebnisse, pädagogische Konsequenzen* (pp. 247-263). Kohlhammer.
- Ronksley-Pavia, M., Barton, G. M., & Pendergast, D. (2019). Multiage education: An exploration of advantages and disadvantages through a systematic review of the literature. *Australian Journal of Teacher Education (Online), 44*(5), 24-41.
- Russell, V. J., Rowe, K. J., & Hill, P. W. (1998). *Effects of Multigrade Classes on Student Progress in Literacy and Numeracy: Quantitative Evidence and Perceptions of Teachers and School Leaders*. Paper presented at the 1998 Annual Conference of the Australian Association for Research in Education
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: examining reasons for acting in two domains. *Journal of personality and social psychology, 57*(5), 749.
- Sewasew, D., & Schroeders, U. (2019). The developmental interplay of academic self-concept and achievement within and across domains among primary school students. *Contemporary Educational Psychology, 58*, 204-212.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of educational research, 46*(3), 407-441.

- Guo, S. & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications*. Sage Publications.
- Spinath, B. (2004). Determinanten von Fähigkeits selbstwahrnehmungen im Grundschulalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36(2), 63-68.
- Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of educational research*, 65(4), 319-381.
- Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data visualisations using the grammar of graphics*. Version, 2(1), 1-189.
- Wigfield, A., & Cambria, J. (2010). Achievement motivation. *The Corsini encyclopedia of psychology*, 1-2.
- Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology*, 110(7), 1005-1025.
- Wolff, F., Nagy, G., Retelsdorf, J., Helm, F., Köller, O., & Möller, J. (2019). The 2I/E model: Integrating temporal comparisons into the internal/external frame of reference model. *Journal of Educational Psychology*, 111(7), 1131.