Käfer, Julia; Kuger, Susanne; Klieme, Eckhard; Kunter, Mareike

The significance of dealing with mistakes for student achievement and motivation. Results of doubly latent multilevel analyses

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
European journal of psychology of education 34 (2019) 4, S. 731-753

The significance of dealing with mistakes for student achievement and motivation: results of doubly-latent multilevel analyses

Julia Käfer [a], Susanne Kuger [b], Eckhard Klieme [a], Mareike Kunter [c]

Abstract

From a constructivist perspective on learning, mistakes are seen as natural elements of learning processes. A supportive and constructive way of dealing with student mistakes has shown to affect students' individual motivation and learning performances in a favorable way. In classroom settings, however, making mistakes is not just a personal but also a social event. Dealing with mistakes should therefore be considered as both student-level and classroom-level characteristic. This study investigates three aspects of dealing with mistakes and their relevance for students' achievement and motivation in English as a foreign language by analyzing the aspects at student and classroom level. The aspects are a) teacher attitude toward mistakes, b) teacher response to student mistakes, and c) students' perception of mistakes as useful for learning. Analyzing data of 5266 students from 427 classes in German secondary schools, the results demonstrate that at student level, all three aspects of dealing with mistakes affect students' individual achievement as well as motivation in English language class. At classroom level, none of the aspects affect student average achievement. Two aspects affect student average motivation at the classroom level, namely a) teacher attitude toward mistakes and c) students' shared perception of the usefulness of mistakes for learning. Our results show that students' individual and shared perception of dealing with mistakes affect students' motivational and cognitive learning outcomes in different ways. Furthermore, our findings underline the relevance of teachers' attitudes as well as students' perceptions concerning mistakes for student learning and motivation in English as a foreign language class.

Keywords: Mistake culture, Error management, Classroom teaching and learning, English as a foreign language, Multilevel analysis

Introduction

Making mistakes is an everyday phenomenon in classroom settings. From a constructivist perspective on learning, mistakes can provide information about the learners' underlying misconceptions or faulty learning processes (Hesketh 1997; Vosniadou and Brewer 1987). Giving the opportunity to reflect on mistakes helps

learners to establish correct mental models and hence to prevent future mistakes (Kapur 2010; Kapur 2014; VanLehn et al. 2003). Therefore, mistakes are seen as natural elements of learning and reflection processes.

Theoretical and empirical evidence indicates that teachers' responses to student mistakes in a supportive and constructive way foster the quantity and quality of students' learning processes (Spychiger et al. 1999). Teachers' dealing with student mistakes affects students' experience with and perception of mistakes as well as their attitudes toward mistakes (Kreutzmann et al. 2014; Santagata 2005; Tulis 2013). Teacher–student interactions concerning mistakes have also been shown to affect students' learning motivation and learning performance (Ames and Archer 1988; Heinze and Reiss 2007; Spychiger et al. 1999; Steuer and Dresel 2015). For instance, a decrease of student motivation might follow from the association of mistakes with negative feelings (Tulis and Fulmer 2013), resulting in attempts to avoid mistakes (Oser and Spychiger 2005).

Following this idea, dealing with mistakes can be seen as an important part of teaching and teaching quality, and should therefore be investigates as a classroom characteristic. Since most research on dealing with mistakes in classrooms investigates students' individual perception as well as their individual activities in mistake situations, evidence on effects of dealing with mistakes as a classroom-level characteristic is scarce and limited to cross-sectional analyses (e.g., Heinze et al. 2012; Kreutzmann et al. 2014; Steuer 2014). The present study therefore targets this research gap and aims to provide evidence on dealing with mistakes at student and classroom level by using a large sample size, two measurement time points, and state-of-the-art methods of analysis.

Theoretical background

Mistake culture

Making mistakes is not just an individual but also a social event (Billett 2012). To describe how the learning environment responds to mistakes, Oser and colleagues introduced the concept of "mistake culture" (e.g., Oser et al. 1999; Oser and Spychiger 2005). In classroom contexts, mistake culture refers to both teachers and students. Accordingly, mistake culture also refers to mistakes that are made by teachers as well as by students. It includes concrete and visible actions in mistake situations as well as mistake-related attitudes promoted by the teacher and shared by students. Usually, the adjectives "positive" and "negative" are used to describe a classroom's mistake culture.

In classrooms that are characterized by a positive mistake culture, teachers encourage their students to reflect on mistakes. Moreover, teachers who promote a positive mistake culture use their knowledge of students'

misconceptions to enhance learning processes. In contrast, a negative mistake culture in class emerges when students try to avoid making mistakes in public because of perhaps being negatively evaluated by the teacher or their peers. In such cases, the potential of learning from mistakes cannot be realized because mistakes are not seen as learning opportunities but potentially harmful for self-related beliefs (Ames and Archer 1988; Oser and Spychiger 2005; Spychiger et al. 2006; Tulis 2013). In classrooms characterized by a negative mistake culture, even teachers try to avoid the occurrence of mistakes; they typically do so by asking questions in such a way that students can hardly give wrong answers (Oser and Spychiger 2005; Spychiger et al. 1999). Oser and colleagues assume that this kind of maladaptive behavior toward mistakes results from an insufficient awareness concerning the relevance of mistakes for the students' knowledge acquisition (see also Santagata 2005; Tulis 2013).

Hence, two basic claims for a supportive and constructive way of dealing with mistakes can be derived from the concept of mistake culture. First, teacher should create a classroom climate that accepts and tolerates student mistakes in learning situations, and second, they should use the learning potential of mistake situations by providing opportunities to reflect on mistakes (see also Heinze and Reiss 2007; Reusser 1999; Spychiger et al. 1999; Türling et al. 2012).

Multidimensionality

The concept of mistake *culture* already indicates that dealing with mistakes is a complex construct. To begin with, different agents are involved in mistake situations (the teacher, the individual student, his/her classmates; Billett 2012; Steuer 2014). Also, dealing with mistakes serves a "social and cognitive dual function" (Klieme 2006, p. 772), meaning that dealing with mistakes affects students' affective and motivational as well as cognitive and behavioral processes (Dresel et al. 2013; Link 2012). Because diverse factors are involved, the construct of dealing with mistakes is assumed and conceptualized to be multidimensional (Heinze and Reiss 2007; Oser and Spychiger 2005; Steuer et al. 2013).

Spychiger and colleagues (1998) were among the first authors who aimed at assessing "dealing with mistakes in class". They designed a student questionnaire (*S-UFS*) with 10 postulated dimensions, which reached popularity in the German-speaking educational research community. Most dimensions of *S-UFS* refer to teacher behavior in mistake situations and to students' individual perception and evaluation of mistakes. In addition, one dimension focuses on how students perceive their teacher's attitude toward mistakes in learning situations. By using exploratory factor analysis, Spychiger et al. (1998) created a short version of the student questionnaire

with three dimensions: 1) teacher behavior in mistake situations, 2) students' use of mistakes as individual learning opportunities, and 3) fear of making mistakes in class (see also Spychiger et al. 2006).

Heinze and colleagues (2012) used an adapted version of this questionnaire and found four dimensions of dealing with mistakes in mathematics. The fourth dimension was identified by splitting the dimension 1) teacher behavior in mistake situations into 1a) cognitive teacher support in using mistakes as learning opportunities, and 1b) affective teacher support in mistake situations. Heinze et al. (2012) based their decision to separate these two dimensions on theoretical reasons, assuming cognitive and affective teacher support to play different roles in mistake situations.

Bearing in mind the different agents who are involved in mistake situations as well as the idea of the social and cognitive dual function of dealing with mistakes, it seems reasonable to assume multidimensionality of the construct. Moreover, we can then assume that different aspects of dealing with mistakes affect students' cognitive and motivational learning outcomes differently (Dresel et al. 2013; Steuer and Dresel 2015).

Empirical findings concerning dealing with mistakes

Domain specificity

Oser and Spychiger (2005) as well as Tulis (2013) reported observational findings that suggest domain-specific differences in teacher responses to students' mistakes. However, most of the empirical studies on dealing with mistakes have focused on mathematics (e.g., Heinze and Reiss 2007; Kreutzmann et al. 2014; Oser and Spychiger 2005; Santagata 2005; Schoy-Lutz 2005; Steuer and Dresel 2015; Steuer et al. 2013; Tulis 2013). Dealing with mistakes in domains other than mathematics has scarcely been studied, for example for German (Tulis 2013; Kreutzmann et al. 2014), history (Oser and Spychiger 2005), physics (Seidel et al. 2003), and economics (Tulis 2013; Mindnich et al. 2008). With regard to English as a foreign language, research in the area of didactics has focused on types and frequency of occurrence of student mistakes as well as on types of teacher reactions (Bohnensteffen 2010). Helmke et al. (2008) presented first empirical evidence on positive relations between a positive mistake culture and students' achievement as well as interest in English as a foreign language. Helmke et al. (2008) used data from the German DESI study (DESI-Konsortium 2008), which we use for secondary analyses in the present study.

Relations to student learning outcomes

A number of studies exist concerning the relationship of dealing with mistakes with affective and respectively motivational outcomes like academic self-efficacy, effort investment, and joy of learning (Kreutzmann et al. 2014), academic self-concept and mastery goal orientation (Steuer et al. 2013), fear of making mistakes and positive learning orientation toward mistakes (Zander et al. 2014; Rach et al. 2013). On the other hand, studies investigating relations to cognitive student learning outcomes like test achievement (Heinze and Reiss 2007; Steuer and Dresel 2015) or grades (Kreutzmann et al. 2014) are scarce. Still, positive correlations between a positive way of dealing with mistakes and student achievement have been found for different subject domains (Steuer and Dresel 2015; Kreutzmann et al. 2014; Helmke et al. 2008). To our knowledge, no study exists which simultaneously takes both outcomes into account.

Individual and contextual characteristics

In most empirical studies, dealing with mistakes was analyzed at the student level investigating students' individual perception and activities concerning dealing with and learning from mistakes. At the classroom level, however, dealing with mistakes has rarely been analyzed (Heinze et al. 2012; Kreutzmann et al. 2014; Steuer et al. 2013; Steuer and Dresel 2015). The body of evidence is therefore small concerning dealing with mistakes as a truly group or classroom characteristic.

As Billett (2012) points out, mistakes have personal and social connotations which influence both the perception of and learning from mistakes. So, since making mistakes is a personal event with individual reaction and perception, it is important to look at the individual student. However, it is just as important to look at students' shared perception of dealing with mistakes. Students' shared perception of the way how mistakes are dealt with in their class can be understood as an indicator of what Oser and colleagues called a classroom's mistake culture. Besides, students' shared perception may affect students' learning outcomes over and above their individual perception (Lüdtke et al. 2009). Therefore, when investigating the significance of dealing with mistakes for student learning, both student-level and classroom-level effects should be considered.

Research aim

Given the existing work in educational contexts (e.g, Spychiger et al. 1998; Steuer et al. 2013), we considered dealing with mistakes to be a multidimensional construct that includes individual as well as contextual components (Billett 2012). The main purpose of our work was therefore to investigate different aspects of

dealing with mistakes and how they affect students' cognitive and motivational learning outcomes when analyzed as both student-level and classroom-level characteristics. Concerning the classroom-level effects, we particularly aimed to examine whether students' shared perception of the different aspects of dealing with mistakes affects students' learning outcomes over and above their individual perception (Lüdtke et al. 2009).

DESI study

To fulfil our research aim, we used data from the German *Assessment of Student Achievements in German and English as a Foreign Language* (DESI) study (DESI-Konsortium 2008). The DESI study was a representative large-scale panel study conducted in almost all secondary school types in Germany. In DESI, ninth-grade students were assessed regarding their language achievement in English as a foreign language and were asked about their learning motivation in this subject domain. Additionally, information on classroom processes including dealing with mistakes was collected. The assessment took place at the beginning and at the end of the 2003/2004 school year (for further information regarding the DESI study design, see Beck et al. 2008).

We decided to use the DESI data even though the study was carried out 15 years ago because, to our best knowledge, there is no recent large-scale study that assessed dealing with mistakes in classroom settings. But also with regards to content, we do not assume that the way of dealing with mistakes in German English language class has notably changed in the last 15 years. At that time, the constructivist perspective on learning was already pursued in classrooms, which also applied to the way how mistakes were dealt with. That is why we assume that findings derived from the DESI data are still interesting for current research and professionals.

At the same time, there were major benefits in using the DESI data for our analyses. Owing to the large sample size, it was possible to use state-of-the-art methods of analysis like multilevel structure equation modelling. In addition, the two measurement time points allowed the control of students' prior knowledge and prior motivation when analyzing the effects of dealing with mistakes on students' learning outcomes.

DESI data have already been used for secondary analyses (e.g., Hochweber and Vieluf 2016; Praetorius et al. 2016; Rjosk et al. 2015) but have not yet been used for our research purpose.

Aspects of dealing with mistakes

In our study, we focused on three aspects of dealing with mistakes which were derived from the DESI student questionnaire: a) teacher attitude toward mistakes, b) teacher response to student mistakes, and c) students' perception of mistakes as useful for learning. The questions were adapted from Spychiger et al. (1998) with the

intention to assess students' perception of their classroom's mistake culture. The three aspects address teacher as well as student perspective in mistake situations but focus the motivational-affective dimension of dealing with mistakes more than the cognitive dimension.

Teacher attitude toward mistakes: The first aspect describes a teacher's tolerant attitude toward students' making mistakes in learning situations (Spychiger 2003; Spychiger et al. 2006). Empirical studies investigating this aspect reveal positive influence on students' perception of mistake situations and their teacher's affective support (Heinze et al. 2012; Heinze and Reiss 2007; Kreutzmann et al. 2014).

Teacher response to student mistakes: Our second aspect refers to the extent to which a teacher provides constructive support in mistake situations, e.g., by giving students a chance to rethink and correct their mistakes. This aspect also refers to the degree to which a teacher's feedback helps students to understand their misconceptions. Research on (mistake-related) feedback confirms that constructive responses on student mistakes support the development of students' motivational (Rakoczy et al. 2008) as well as cognitive (Heinze and Reiss 2007) learning outcomes.

Students' perception of mistakes as useful for learning: The last aspect concerns students' perception of mistakes as learning opportunities and describes the extent to which students' evaluate the function and usability of mistakes as being useful for learning. Findings on students' perceived usefulness of feedback, including feedback on mistakes, reveal that students, who perceive the provided feedback as useful show higher interest (Rakoczy et al. 2013), self-efficacy (Rakoczy et al. 2018), joy of learning (Kreutzmann et al. 2014), and even achievement (Harks et al. 2014).

Notably, evidence on these three aspects mostly applies to student-level effects representing students' individual perception of their teacher's as well as their own dealing with mistakes. The evidence on student-level effects cannot be directly transferred to classroom-level effects, because the meaning of constructs might change after aggregating variables at the cluster-level (van de Vijver et al. 2008). As mentioned above, so far there is little research on effects of dealing with mistakes as a classroom-level characteristic (Heinze et al. 2012; Kreutzmann et al. 2014; Steuer 2014; Steuer et al. 2013; Steuer and Dresel 2015).

Research questions and hypotheses

Building on the stated research desiderata and the literature review concerning the three aspects of dealing with mistakes, the following research questions guided our study.

Do the three aspects of dealing with mistakes a) teacher attitude toward mistakes, b) teacher response to student mistakes, and c) students' perception of mistakes as useful for learning affect 1) student achievement and 2) motivation in English as a foreign language when considering the aspects as both individual-level and classroom-level characteristics, and after controlling for prior knowledge and prior motivation, respectively? Concerning the investigation of the classroom-level effects, we are particularly interested whether students' shared perception of the aspects affects students' average achievement and motivation in a classroom over and above their individual perception.

In view of prior research (e.g., Heinze et al. 2012; Heinze and Reiss 2007; Kreutzmann et al. 2014; Rakoczy et al. 2008), we hypothesize that on the individual level all three aspects affect student achievement (Hypotheses 1a – 1c) as well as motivation (Hypotheses 2a – 2c) positively. On the classroom level, however, we do not formulate concrete hypotheses due to the small body of evidence on dealing with mistakes as a classroom-level characteristic.

Method

Sample

The sample drawn for the DESI study consists of whole classrooms and is representative for ninth-grade students attending secondary school (i.e., academic track, intermediate track, and lower track secondary schools) in Germany. For our analyses, we used a sample of 5266 students out of 427 classes who were asked to answer questions about classroom processes in their English language class. Of the 5266 students 53% were female, 48% attended the academic track, 36% the intermediate track, and 16% the lower track. The students' mean age was 14.8 years (SD=0.73, range=13-18) at the beginning of the study. As a general rule, German ninth-grade students are in their fifth year of learning English as a foreign language.

Measures

Student achievement: To measure student achievement in English language, a C test (Harsch and Schröder 2007) and a listening comprehension test (Nold and Rossa 2007) were administered at the beginning (T1) and at the end (T2) of the school year.

C tests are written tests that measure overall language abilities in a foreign language by asking test takers to reconstruct the original message from incomplete texts (Grotjahn 2002). The C test used in the DESI study consisted of eight text passages with each 25 gaps. The listening comprehension test included 16 short dialogs and four longer talks similar to a radio report. A rotated booklet design was used to ensure that students worked on different items at T1 and T2 for both the C test and the listening comprehension test.

The scores were scaled based on a generalized Rasch model (Adams and Wu 2007) using ConQuest (Wu et al. 1998). For more detailed information regarding the scaling process executed in DESI, see Hartig et al. (2008).

Weighted Likelihood Estimates (WLE; Warm 1989) were computed and used as achievement scores for both dimensions of student achievement at T1 and T2. For the C test, the estimated reliability (EAP/PV) was 0.93 at T1 and 0.87 at T2; for the listening comprehension test, the reliability was 0.70 at T1 and 0.66 at T2.

We checked the indices of intra-class correlations, ICC1 and ICC2, for both tests and measurement time points. Indices of intra-class correlations are used to assess whether aggregated individual-level assessments (or ratings) are reliable indicators of class-level constructs (e.g., Lüdtke et al. 2006; Lüdtke et al. 2009). ICC1 refers to individual-level assessments (or ratings) and indicate the proportion of variance in an observed variable that can be explained by differences between classes. ICC2, in contrast, is a measure of reliability of class-mean assessments (or ratings). In our study, the indices indicate substantial differences between classes (ICC1) and high reliability on the classroom level (ICC2) for the C test (T1: ICC1=0.66, ICC2=0.96; T2: ICC1=0.64, ICC2=0.95) as well as the listening comprehension test (T1: ICC1=0.48, ICC2=0.91; T2: ICC1=0.54, ICC2=0.93).

Student motivation: Motivation in English language was assessed at the beginning (T1) and at the end (T2) of the school year by using a five-item scale ranging from 1 (=strongly disagree) to 4 (=strongly agree). The items (e.g., "English is fun." [1]) were adapted from Rheinberg and Wendland's (2001) Potsdam Motivation Inventory (PMI).

Internal consistency of the scale was good (Cronbach's α=0.81 at T1 and α=0.83 at T2). The scale's mean and standard deviation were *M*=2.64 and *SD*=0.71 at T1, and *M*=2.55 and *SD*=0.77 at T2. According to the intra-class correlations, there was a moderate amount of variability among classes and a sufficiently high

[1] The items were translated into English by the authors.

reliability on the classroom level for both measurement time points (T1: ICC1=0.15, ICC2=0.65; T2: ICC1=0.15, ICC2=0.66).

Aspects of dealing with mistakes: Dealing with mistakes was assessed via student ratings at the end of the school year (T2). Among other classroom processes, students rated a series of items on a four-point scale ranging from 1 (=strongly disagree) to 4 (=strongly agree) how they perceived the mistake culture in their English class. The scale was adapted from *S-UFS* (Spychiger et al. 1998). For our analyses, we used three aspects of dealing with mistakes, which are a) teacher attitude toward mistakes (three items), b) teacher response to student mistakes (four items), and c) students' perception of mistakes as useful for learning (two items). In Table 1 item formulations, internal consistencies, means and standard deviations of the aspects are summed up.

Insert Table 1 here

Indices of intra-class correlations indicate moderate differences between classes and sufficiently high agreement within classes for all three aspects of dealing with mistakes (teacher attitude toward mistakes: ICC1=0.15, ICC2=0.65; teacher response to student mistakes: ICC1=0.15, ICC2=0.66; students' perception of mistakes as useful for learning: ICC1=0.08, ICC2=0.48). According to the indices, the aspects can be analyzed as classroom-level characteristics representing students' shared perception of dealing with mistakes (Lüdtke et al. 2009; Stapleton et al. 2016).

Analyses

In order to investigate the influence of the three aspects of dealing with mistakes on student achievement and motivation, we applied doubly latent multilevel structural equation modeling (ML-SEM) proposed by Marsh and colleagues (2009; 2012; also Lüdtke et al. 2008; Lüdtke et al. 2011). ML-SEM provides a possibility to estimate the structural relations among unobserved latent variables of a hypothesized model with a simultaneous consideration of the multi-level data structure. Since ML-SEM controls for both measurement error and sampling error, this method is considered to be more appropriate to investigate classroom-level characteristics than traditional multi-level analyses using manifest variables and manifest aggregation (Lüdtke et al. 2011; Marsh et al. 2009).

In our analyses, we used the three aspects of dealing with mistakes as separate predictor variables to avoid potential confounding effects. Each aspect was represented by multiple indicators. As we considered dealing with mistakes to be a student-level construct, representing students' individual perception, as well as a classroom-level construct, representing students' shared perception, the three aspects were modelled at both levels simultaneously (see Figure 1).

Student achievement and motivation in English language at the end of the school year (T2) served as outcome variables and were also represented at both levels by using multiple indicators. For reasons of model parsimony, we controlled for only one covariate in each model. This covariate was either prior knowledge (T1 achievement) or prior motivation (T1 motivation) depending on the outcome variable.

All in all, we modelled six ML-SEM models (2 outcomes x 3 aspects of dealing with mistakes). When analyzing the classroom-level effects of dealing with mistakes, we controlled for the corresponding student-level effects. The presented classroom-level coefficients are therefore contextual effects representing the effects of students' shared perception over and above students' individual perception (Marsh et al. 2012). In Figure 1, this is operationalized as the difference between the classroom-level and the student-level effects (shown in Figure 1 as hexagons with dashed lines).

Insert Figure 1 here

All models were estimated in Mplus 7.11 (Muthén and Muthén 1998–2013), using maximum likelihood estimation with robust standard errors. Sampling weights were used to adjust for unequal sampling probabilities. All variables were group-mean centered due to the decomposition of effects into student-level and classroom-level components, which is implicitly done in all doubly latent ML-SEM models estimated in Mplus (Marsh et al. 2009; Marsh et al. 2012).

The amount of missing data ranged from 5% for prior achievement in the C test to 17% for student prior motivation. To deal with missing data, we used the full information maximum likelihood (FIML) approach incorporated in Mplus.

Results

Relations among latent variables

To better understand the results of the subsequent ML-SEM models, it is useful first to look at the pattern of bivariate correlations. Table 2 provides an overview of the correlations among the latent variables at student and classroom level.

At the student level, test-retest correlations for achievement and motivation were high for both constructs. High correlations could also be found between the three aspects of dealing with mistakes. We therefore analyzed the aspects in separate ML-SEM models. Besides, the correlations of the aspects with motivation were higher than with achievement – this applied to both measurement time points.

At the classroom level, the correlation pattern was almost similar. Correlations between the three aspects of dealing with mistakes were high. Correlations between the aspects and motivation were higher than between the aspects and achievement. However, at the classroom level the correlation between teacher response to student mistakes and student achievement was the lowest at both time points.

Insert Table 2 here

Latent multilevel modeling

Before analyzing the six ML-SEM models, we tested measurement invariance across the two levels via multilevel confirmatory factor analyses (ML-CFA; Muthén and Asparouhov 2011). This was mandatory because we treated the aspects of dealing with mistakes as *configural constructs* (Stapleton et al. 2016). A configural construct is defined as an "aggregate of the measurements of individuals who comprise the cluster" (Stapleton et al. 2016, p. 496). Therefore, this modeling approach implies an assumption that unstandardized classroom-level loadings are fixed to be the same as student-level loadings (see also Marsh et al. 2012; Morin et al. 2014).

We also tested for measurement invariance across the two measurement time points in order to reduce the amount of model parameters in our main specification ML-SEM models.

Goodness of fit was assessed with the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA), as operationalized in Mplus by applying the MLR estimator (Muthén & Muthén, 1998-2013). The relative fit indices show that there was no substantial decrement in fit for any of the tested invariance constraints (for detailed information, see Appendix A). Here, we used

cutoffs recommended by Rutkowski and Svetina (2014) like the ΔCFA around -0.020 or ΔRMSEA criterion of 0.030.

Hence, all following results are based upon the invariance model in which factor loadings were constrained to be invariant across levels and over time. Substantially, this implies that the construct's meaning at T1 and at T2 as well as on student level and on classroom level can be considered equal.

Regarding the factor loadings in our main specification ML-SEM models, results showed substantial and significant factor loadings for all multiple-indicator constructs at student and classroom level. At the student level, the standardized factor loadings varied from 0.48 to 0.80 for teacher attitude toward mistakes, from 0.50 to 0.76 for teacher response to mistakes, and from 0.55 to 0.75 for students' perception of mistakes as useful learning opportunities. At the classroom level, the standardized factor loadings were very high, as is typically the case (Marsh et al. 2012). For teacher attitude toward mistakes, the loadings varied from 0.61 to 1.07, for teacher response to mistakes from 0.72 to 1.01, and for students' perception of mistakes as learning opportunities from 0.84 to 1.04.

Table 3 shows the results of the six ML-SEM models. In Models 1a – 1c, achievement in English as a foreign language served as the outcome variable. The results revealed that at the student level all aspects had a significant positive relation with student individual achievement. At the classroom level, however, the opposite pattern emerged; none of the aspects affect student average achievement significantly.

In Models 2a – 2c, motivation in English language was the outcome. Similar to the results on achievement, all three aspects showed significant positive relations at student level. Moreover, at the classroom level, students' shared perception of teacher attitude toward mistakes as well as of the usefulness of mistakes for learning showed significant positive relations with student average motivation.

According to these results, students' individual achievement as well as motivation was higher when they perceived their teacher's attitude to be tolerant and mistake-friendly and their teacher's response to mistakes as being positive and constructive. In addition, students' individual achievement and motivation was also higher when students themselves perceived mistakes to be useful for learning.

Concerning the classroom level, students' average achievement was not influenced by their shared perception of how mistakes are dealt with in their English language class. Rather, students' average motivation was higher in classrooms in which the teacher shows a positive attitude toward mistakes, and in which students have a shared understanding of mistakes as being useful for learning.

Insert Table 3 here

Discussion

The purpose of the present work was to investigate how different aspects of dealing with mistakes affect 1) student achievement and 2) motivation in English as a foreign language when the aspects are analyzed as both student-level and classroom-level characteristics. Conducting a secondary analysis of data from the German DESI study (DESI-Konsortium 2008), we focused on three aspects of dealing with mistakes, which were assessed via student questionnaire: a) teacher attitude toward mistakes, b) teacher response to student mistakes, and c) students' perception of mistakes as useful for learning. To analyze student-level and classroom-level effects simultaneously, we applied the doubly latent multilevel structure equation modeling approach (Marsh et al. 2012).

By analyzing dealing with mistakes not only at student level, representing students' individual perception, but also at classroom level, representing students' shared perception, our study contributes to a better understanding of the role dealing with mistakes plays for student learning. Students' shared perception of the way how mistakes are dealt with in their classroom refers to Oser and colleagues' concept of "mistake culture" (e.g., Oser et al. 1999), and may affect students' learning outcomes over and above their individual perception (Lüdtke et al. 2009). In our study, we therefore controlled for the student-level effects when investigating the classroom-level effects of the three aspects (contextual effects; see Marsh et al. 2012).

Student-level effects

As hypothesized, on student level, we found positive relations with students' individual achievement (Hypotheses 1a – 1c) as well as individual motivation (Hypotheses 2a – 2c) for all three aspects of dealing with mistakes. Thus, our findings correspond with previous research findings (Kreutzmann et al. 2014; Rakoczy et al. 2008; Rakoczy et al. 2018; see also Oser and Spychiger 2005), while delivering new insights regarding the significance of dealing with mistakes for students' individual achievement and motivation. For instance, prior research showed that students who perceive their teacher's attitude as tolerant and mistake-friendly report a positive perception of their teacher's affective support in mistake situations and less anxiety to make mistakes (Heinze and Reiss 2007; Rach et al. 2013). Our results go further by showing that in English class students' individual perception of their teacher's mistake-friendly attitude fosters their individual motivation and even achievement.

Classroom-level effects

With regard to the classroom level, we found an ambiguous pattern of results. None of the aspects affect students' average achievement. Students' average motivation, however, was positively related with two aspects, namely students' shared perception of their teacher's mistake-related attitude and students' shared perception of mistakes as useful for learning.

According to research on teaching quality, supportive climate, as one of three generic quality dimensions of classroom processes (Klieme 2006; Klieme et al. 2006; see also Pianta and Hamre 2009), affects student motivation more than students' cognitive learning outcomes (Klieme et al. 2009; Lipowsky et al. 2009). An appropriate way of dealing with student mistakes can be considered as part of the supportive climate (Helmke 2006; Seidel and Prenzel 2004). A supportive climate in turn fosters students' learning motivation. This and the fact that all aspects of dealing with mistakes investigated in our study referred to the motivational-affective dimension of dealing with mistakes might be the reason that we found significant relations of dealing with mistakes and students' average motivation but not with students' average achievement. Admittedly, prior knowledge (T1 achievement) was highly predictive for student achievement at T2 leaving very little variance to be predicted by other variables.

With regard to the positive classroom-level effect of a teacher's mistake-related attitude on student motivation, evidence of prior research is unclear. Kreutzmann et al. (2014) used an adapted version of the student questionnaire from Spychiger et al. (2006), but found no effects of a teacher's mistakes friendliness, considered as classroom-level characteristic, on students' motivational outcomes like academic self-efficacy, effort investment, and joy of learning. Even though Kreutzmann et al. (2014) used a similar instrument to assess dealing with mistakes, our findings cannot be easily compared. While Kreutzmann et al. (2014) investigated 421 primary school students in mathematics using a cross-sectional study design, we investigated the development of 5266 secondary school students in English as a foreign language using a panel study design with two waves of data collection. The major differences in study design do not allow for a deduction concerning the significance of teacher mistake-related attitude for student learning. To clarify the picture, further evidence is needed.

The second classroom-level effect that turned out to be significant for students' average motivation refers to students' shared perception regarding the usefulness of mistakes. Students' shared perception of the usefulness can be interpreted as an indicator of the class's attitude toward the function and usability of mistakes as learning opportunities, but does not necessary concern the concrete *use* of mistakes. Research on feedback has revealed that the functional role of feedback depends on learners' perception and interpretation of the feedback (Black and

Wiliam 2012). Therefore, more empirical evidence on students' perception of usefulness and its effects on student learning can be found in the research field on feedback rather than in research on dealing with mistakes. Rakoczy et al. (2008), for example, examined students' shared perception regarding the usefulness of their teachers' feedback, including feedback on mistakes. They found positive effects on the development of students' self-efficacy at the classroom level. Similarly, we showed that students' shared perception affects students' motivational learning outcomes at the classroom level. Moreover, by identifying the importance of students' shared perception regarding the usefulness of mistakes for students' average motivation, our findings complement the research on dealing with mistakes as a classroom-level characteristic.

Teacher response to student mistakes in turn has shown to be non-significant, so it seems that students' shared perception of their teacher's constructive and supportive responses to their mistakes does not influence their average motivation in class over and above their individual perception. One reason might be that a teacher's responses are adjusted according to the particular mistake as well as to the student's ability level (Brookhart 2008; Rakoczy et al. 2013). Therefore, feedback that helps one student to understand and overcome his/her mistake is not necessarily helpful for all the other students in class. Our results support this assumption by showing that students' individual perception of this aspect positively affects students' individual motivation, and even achievement. Students' shared perception, however, neither affects students' average motivation nor achievement.

Another reason might be that a teacher's feedback on mistakes is not necessarily associated with positive feelings (Oser and Spychiger 2005; Spychiger et al. 2006; Tulis and Fulmer 2013). Students who get feedback on their mistakes may think that they are not smart enough to do it right. This might particularly apply to underachieving students. So, if feedback on mistakes is negatively related to motivation in the group of underachievers but is positively related in the group of high achievers, this can explain why on average we found a null effect. Therefore, further research on the heterogeneity of relations is needed to clarify this issue.

Limitations

Important limitations of our study should be addressed. First of all, students were asked to rate how mistakes are dealt with in their English language class but they were not asked what they understand by the term *mistake*. Students could think about wrong grammar or pronunciation, but also about not knowing the correct answer. Hence, we investigated students' perception of dealing with mistakes without knowing what they considered as mistakes.

Second, our study focused on dealing with mistakes in English as a foreign language. In this specific subject domain, students are expected to be prone to make mistakes, especially on pronunciation and grammar (Bohnensteffen 2010). Oser and Spychiger (2005) as well as Tulis (2013) report domain specific differences in dealing with mistakes. Thus, it is questionable whether our findings can be generalized to other subject domains. Another limitation is that the three investigated aspects focused on the motivational-affective dimension of dealing with mistakes. As dealing with mistakes serves a "social and cognitive dual function" (Klieme 2006, p. 772), there is a cognitive dimension that is instructionally relevant as well (Deppe 2017). In our study, the cognitive dimension of dealing with mistakes was not investigated because it was not considered in the student questionnaires.

At last, it should be mentioned that information on teacher attitude toward mistakes was based on students' reports. Still, since our findings reveal that students' perception of their teacher's mistake-related attitude is important for students' individual as well as average motivation, this aspect of dealing with mistakes should be further investigated. We suggest, however, to use teacher self-reports when investigating teachers' mistake-related concepts and beliefs (e.g., Santagata 2004). Ideally they are used in addition to the student perspective, thus allowing deeper insights into the complexity of dealing with mistakes and its significance for student learning.

Conclusion

The present study supports the significance of dealing with mistakes for student achievement and motivation in classroom settings. Moreover, dealing with mistakes should be considered as both individual-level and classroom-level characteristic. Our results show that students' individual and shared perception of dealing with mistakes affect students' motivational and cognitive learning outcomes in different ways. Especially regarding motivation, students' shared perception revealed effects over and above their individual perception. Hence, as a direction for future research, it can be said that taking a differentiated view on the different levels of analysis seems to be appropriate when investigating the role dealing with mistakes plays for student learning.

Furthermore, we can conclude that it is important for both research and practice to focus on teachers' mistake-related attitudes as well as on students' perception of usefulness of mistakes. Training programs addressing the relevance of mistakes for student knowledge acquisition (e.g., Heinze and Reiss 2007) could be used for interventions or teachers' professional development. Based on our findings, we suggest that helping

teachers and students to change their perception toward mistakes in a positive way should be beneficial to students' individual achievement and motivation, as well as for students' average motivation in class.

Funding information: The present study is a result of the Thin Slices project, which is collaboratively conducted by researchers at the DIPF | Leibniz Institute for Research and Information in Education and the Goethe University Frankfurt. The preparation of this paper was supported by grants from the German Federal Ministry of Education and Research (BMBF, grand number 01LSA1501B).

Appendix

Insert Table 4 here

References

Adams, R. J., and Wu, M. L. (2007). The mixed-coefficients logit model: A generalized form of the Rasch model. In M. v. Davier and C. H. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications* (pp. 57–75). New York: Springer.

Ames, C., and Archer, J. (1988). Achievement Goals in the Classroom: Students' Learning Strategies and Motivation Processes. *Journal of Educational Psychology*, *80*(3), 260–267. https://doi.org/10.1037/0022-0663.80.3.260.

Beck, B., Bundt, S., and Gomolka, J. (2008). Ziele und Anlage der DESI-Studie [Objectives and Layout of the DESI Study]. In DESI-Konsortium (Ed.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (pp. 11–25). Weinheim: Beltz.

Billett, S. (2012). Errors and Learning from Errors at Work. In J. Bauer and C. Harteis (Eds.), *Professional and Practice-based Learning: Vol. 6. Human Fallibility: The Ambiguity of Errors for Work and Learning* (pp. 17–32). Dordrecht: Springer.

Black, P., and Wiliam, D. (2012). Assessment for Learning in the Classroom. In J. Gardner (Ed.), *Assessment and Learning* (pp. 11–32). London: SAGE Publications Ltd. https://doi.org/10.4135/9781446250808.n2.

Bohnensteffen, M. (2010). *Fehler-Korrektur: Lehrer- und lernerbezogene Untersuchungen zur Fehlerdidaktik im Englischunterricht der Sekundarstufe II* [Mistake Correction: Teacher and Learner-Related Studies on Mistake Didactics in English at Upper Secondary School]. *Europäische Hochschulschriften. Reihe XI, Pädagogik: Bd. 1001*. Frankfurt am Main: Lang.

Brookhart, S. M. (2008). *How to give Effective Feedback to Your Students*. Alexandria, VA: Association for Supervision and Curriculum Development.

Deppe, M. (2017). *Fehler als Stationen im Lernprozess: Eine kognitionswissenschaftliche Untersuchung am Beispiel Rechnungswesen* [Mistakes as Stations in the Learning Process: A Cognitive Science Study on the Example of Financial Accounting]. *Wirtschaft - Beruf - Ethik*.

DESI-Konsortium (Ed.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* [Teaching and Acquisition of Competencies in German and English: Results of the DESI Study]. Weinheim: Beltz.

Dresel, M., Schober, B., Ziegler, A., Grassinger, R., and Steuer, G. (2013). Affektiv-motivational adaptive und handlungsadaptive Reaktionen auf Fehler im Lernprozess [Affective-Motivational Adaptive and Action-Adaptive Reactions on Mistakes in Learning Processes]. *Zeitschrift für pädagogische Psychologie*, *27*(4), 255–271. https://doi.org/10.1024/1010-0652/a000111.

Grotjahn, R. (2002). Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen. [The C-Test: Theoretical Groundwork and Practical Applications]* (pp. 211–225). Bochum: AKS.

Harks, B., Rakoczy, K., Hattie, J., Besser, M., and Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*, *34*(3), 269–290. https://doi.org/10.1080/01443410.2013.785384.

Harsch, C., and Schröder, K. (2007). Textrekonstruktion: C-Test. In B. Beck and E. Klieme (Eds.), *Sprachliche Kompetenzen: Konzepte und Messung; DESI-Studie (Deutsch-Englisch-Schülerleistungen-International). [Language Competencies: Concepts and Evaluation; DESI-Study (Assessment of Student Achievement in German and English as a Foreign Language)]* (pp. 212–225). Weinheim: Beltz.

Hartig, J., Jude, N., and Wagner, W. (2008). Methodische Grundlagen der Messung und Erklärung sprachlicher Kompetenzen [Method Foundations of Measuring and Explaining Language Competencies]. In DESI-Konsortium (Ed.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (pp. 34–54). Weinheim: Beltz.

Heinze, A., and Reiss, K. (2007). Mistake-handling activities in the mathematics classroom: Effects of an in-service teacher training on students' performance in geometry. In H. L. Chick and J. L. Vincent (Eds.), *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education* (pp. 9–16).

Heinze, A., Ufer, S., Rach, S., and Reiss, K. (2012). The student perspective on dealing with errors in mathematics class. In E. Wuttke and J. Seifried (Eds.), *Research in vocational education: Vol. 1. Learning from Errors at School and at Work* (pp. 65–80). Opladen: Budrich.

Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Klieme, E., Nold, G., and Schröder, K. (2008). Wirksamkeit des Englischunterrichts [Efficacy of English Language Education]. In DESI-Konsortium (Ed.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (pp. 382–397). Weinheim: Beltz.

Helmke, A. (2006). Was wissen wir über guten Unterricht? Über die Notwendigkeit der Rückbesinnung auf den Unterricht als dem „Kerngeschäft" der Schule [What Do We Know About Good Teaching? About the Necessity of Refocusing on Teaching as the "core business" of School]. *Pädagogik und Schulalltag*, *58*, 42–45.

Hesketh, B. (1997). Dilemmas in Training for Transfer and Retention. *Applied Psychology*, *46*(4), 317–339. https://doi.org/10.1111/j.1464-0597.1997.tb01234.x.

Hochweber, J., and Vieluf, S. (2016). Gender differences in reading achievement and enjoyment of reading: The role of perceived teaching quality. *The Journal of Educational Research*, *111*(3), 268–283. https://doi.org/10.1080/00220671.2016.1253536.

Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde [Empirical Educational Research: Recent Developments, Theoretical Basis and Domain-Specific Results]. *Zeitschrift für Pädagogik*, *52*(6), 765–773.

Klieme, E., Lipowsky, F., Rakoczy, K., and Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht: Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras" [Quality Dimensions and Efficacy of Mathematical Education. Theoretical Basis and Selected Results of the Project "Pythagoras"]. In M. Prenzel (Ed.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms: BIQUA* (pp. 127–146). Münster: Waxmann.

Klieme, E., Pauli, C., and Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In J. Tomáš and T. Seidel (Eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom* (pp. 137–160). Münster: Waxmann.

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science 38*(6), 523–550. https://doi.org/10.1007/s11251-009-9093-x.

Kapur, M. (2012). Productive Failure in Learning Math. *Cognitive Science 38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107.

Kreutzmann, M., Zander, L., and Hannover, B. (2014). Versuch macht kluch/g?! Der Umgang mit Fehlern auf Klassen- und Individualebene.: Zusammenhänge mit Selbstwirksamkeit, Anstrengungsbereitschaft und Lernfreude von Schülerinnen und Schülern [Learning by Doing? Managing Mistakes on the Class and

Individual Level: Interrelations with Students' Self-efficacy, Effort Investment, and Joy of Learning]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *46*(2), 101–113. https://doi.org/10.1026/0049-8637/a000103.

Link, M. (2012). Learning from errors: Strategies for accounting lessons - Outline of a research project. In E. Wuttke and J. Seifried (Eds.), *Research in vocational education: Vol. 1. Learning from Errors at School and at Work* (pp. 49–64). Opladen: Budrich.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., and Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, *19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001.

Lüdtke, O., Marsh, H. W., Robitzsch, A., and Trautwein, U. (2011). A 2x2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error-correction models. *Psychological Methods*, *16*, 444–467.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., and Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229.

Lüdtke, O., Robitzsch, A., Trautwein, U., and Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modelling. *Contemporary Educational Psychology*, *34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001.

Lüdtke, O., Trautwein, U., Kunter, M., and Baumert, J. (2006). Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen [Analysis of Learning Environments: Approaches to Determining Reliability and Consistency with Student Perceptions]. *Zeitschrift für pädagogische Psychologie*, *20*(1/2), 85–96.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., and Köller, O. (2012). Classroom Climate and Contextual Effects: Conceptual and Methodological Issues in the Evaluation of Group-Level Effects. *Educational Psychologist*, *47*(2), 106–124.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., and Nagengast, B. (2009). Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. *Multivariate Behavioral Research*, *44*, 764–802. https://doi.org/10.1080/00273170903333665.

Mindnich, A., Wuttke, E., and Seifried, J. (2008). Aus Fehlern wird man klug? Eine Pilotstudie zur Typisierung von Fehlern und Fehlersituationen [Learning from Mistakes? A Pilot Study on the Typification of Mistakes and Mistake Situations]. In E.-M. Lankes (Ed.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (pp. 153–163). Münster: Waxmann.

Morin, A. J. S., Marsh, H. W., Nagengast, B., and Scalas, L. F. (2014). Doubly Latent Multilevel Analyses of Classroom Climate: An Illustration. *The Journal of Experimental Education, 82*(2), 143–167. https://doi.org/10.1080/00220973.2013.769412.

Muthén, B. O., and Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox and J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). New York: Taylor and Francis.

Muthén, L. K., and Muthén, B. O. (1998–2013). *Mplus: Statistical analysis with latent variables: User's guide. Version 7.11*.

Nold, G., and Rossa, H. (2007). Hörverstehen. In B. Beck and E. Klieme (Eds.), *Sprachliche Kompetenzen: Konzepte und Messung; DESI-Studie (Deutsch-Englisch-Schülerleistungen-International). [Language Competencies: Concepts and Evaluation; DESI-Study (Assessment of Student Achievement in German and English as a Foreign Language)]* (pp. 174–192). Weinheim: Beltz.

Oser, F., Hascher, T., and Spychiger, M. B. (1999). Lernen aus Fehlern: Zur Psychologie des „negativen" Wissens [Learning from Mistakes: On the Psychology of „Negative" Knowledge]. In W. Althof (Ed.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlaß des 60. Geburtstags von Fritz Oser* (pp. 11–41). Wiesbaden: Springer.

Oser, F., and Spychiger, M. B. (2005). *Lernen ist schmerzhaft: Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur* [Learning is Painful: On the Theory of Negative Knowledge and the Practice of Mistake Culture]. *Beltz-Pädagogik*. Weinheim: Beltz.

Pianta, R. C., and Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, *38*(2), 109–119.

Praetorius, A.-K., Vieluf, S., Saß, S., Bernholt, A., and Klieme, E. (2016). The same in German as in English? Investigating the subject-specificity of teaching quality. *Zeitschrift für Erziehungswissenschaft*, *19*(1), 191–209. https://doi.org/10.1007/s11618-015-0660-4.

Rach, S., Ufer, S., and Heinze, A. (2013). Learning from errors: effects of teachers' training on students' attitudes towards and their individual use of errors. *PNA*, *8*(1), 21–30.

Rakoczy, K., Harks, B., Klieme, E., Blum, W., and Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, *27*, 63–73. https://doi.org/10.1016/j.learninstruc.2013.03.002.

Rakoczy, K., Klieme, E., Bürgermeister, A., and Harks, B. (2008). The Interplay Between Student Evaluation and Instruction: Grading and Feedback in Mathematics Classrooms. *Zeitschrift für Psychologie*, *216*(2), 111–124. https://doi.org/10.1027/0044-3409.216.2.111.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., and Besser, M. (2018). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2018.01.004.

Reusser, K. (1999). Schülerfehler – die Rückseite des Spiegels [Student Mistakes – the Other Side of the Mirror]. In W. Althof (Ed.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlaß des 60. Geburtstags von Fritz Oser* (pp. 203–231). Wiesbaden: Springer.

Rheinberg, F., and Wendland, M. (2001). *Veränderung der Lernmotivation in Mathematik und Physik: Eine Komponentenanalyse und der Einfluss elterlicher sowie schulischer Kontextfaktoren*. DFG-Bericht zum Projekt „Förderung von Motivationskomponenten" [Changes of Motivation for Learning in Mathematics and Physical Science: A Component Analysis and the Influence of Parental and School Context Factors. DFG-Report on the Project "Promoting Motivational Components"].

Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., and Stanat, P. (2015). Classroom composition and language minority students' motivation in language lessons. *Journal of Educational Psychology*, *107*(4), 1171–1185. https://doi.org/10.1037/edu0000035.

Rutkowski, L., and Svetina, D. (2014). Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys. *Educational and Psychological Measurement, 74*(1), 31–57. https://doi.org/10.1177/0013164413498257.

Santagata, R. (2004). "Are you joking or are you sleeping?" Cultural beliefs and practices in Italian and U.S. teachers' mistake-handling strategies. *Linguistics and Education*, *15*(1-2), 141–164. https://doi.org/10.1016/j.linged.2004.12.002.

Santagata, R. (2005). Practices and beliefs in mistake-handling activities: A video study of Italian and US mathematics lessons. *Teaching and Teacher Education*, *21*(5), 491–508. https://doi.org/10.1016/j.tate.2005.03.004.

Schoy-Lutz, M. (2005). *Fehlerkultur im Mathematikunterricht: Theoretische Grundlegung und evaluierte unterrichtspraktische Erprobung anhand der Unterrichtseinheit „Einführung in die Satzgruppe des Pythagoras"*. [Mistake Culture in Mathematics: Theoretical Basis and Evaluated Practical Trial in Class for the Teaching Unit "Introduction to the Pythagorean Theorem"]. *Texte zur mathematischen Forschung und Lehre: Vol. 39*. Hildesheim: Franzbecker.

Seidel, T., Kobarg, M., and Rimmele, R. (2003). Aufbereitung der Videodaten [Preparation of Video Data]. In T. Seidel, M. Prenzel, R. Duit, and M. Lehrke (Eds.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht"*. Kiel: IPN.

Seidel, T., and Prenzel, M. (2004). Muster unterrichtlicher Aktivitäten im Physikunterricht [Patterns of Teaching Activities in Physics]. In J. Doll and M. Prenzel (Eds.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung*. Münster: Waxmann.

Spychiger, M. B. (2003). Fehler als Fenster auf den Lernprozess: Zur Entwicklung einer Fehlerkultur in der Praxisausbildung [Mistakes as Windows to the Learning Process: On the Development of a Mistake Culture in Teaching Practices]. *Journal für LehrerInnenbildung*, *3*(2), 31–38.

Spychiger, M. B., Kuster, R., and Oser, F. (2006). Dimensionen von Fehlerkultur in der Schule und deren Messung: Der Schülerfragebogen zur Fehlerkultur im Unterricht für Mittel- und Oberstufe [Dimensions of Mistake Culture in School and their Measurement: The Student Questionnaire on Mistake Culture in Secondary School Classes]. *Schweizerische Zeitschrift für Bildungswissenschaften*, *28*(1), 87–110.

Spychiger, M. B., Mahler, F., Hascher, T., and Oser, F. (1998). *Fehlerkultur aus der Sicht von Schülerinnen und Schülern: Der Fragebogen S-UFS*. Entwicklung und erste Ergebnisse [Mistake Culture as Viewed by Students: The S-UFS Questionnaire. Development and First Results]. *Lernen Menschen aus Fehlern? Zur Entwicklung einer Fehlerkultur in der Schule: Vol. 4*. Freiburg, Schweiz: Pädagogisches Institut der Universität Freiburg.

Spychiger, M. B., Oser, F., Hascher, T., and Mahler, F. (1999). Zur Entwicklung einer Fehlerkultur in der Schule [On the Development of a Mistake Culture in School]. In W. Althof (Ed.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlaß des 60. Geburtstags von Fritz Oser* (pp. 43–70). Wiesbaden: Springer.

Stapleton, L. M., Yang, J. S., and Hancock, G. R. (2016). Construct Meaning in Multilevel Settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. https://doi.org/10.3102/1076998616646200.

Steuer, G. (2014). *Fehlerklima in der Klasse: Zum Umgang mit Fehlern im Mathematikunterricht* [Mistake Climate in Class: About Dealing with Mistakes in Mathematics]. Dordrecht: Springer.

Steuer, G., and Dresel, M. (2015). A constructive error climate as an element of effective learning environments. *Psychological Test and Assessment Modeling*. (57), 262–275.

Steuer, G., Rosentritt-Brunn, G., and Dresel, M. (2013). Dealing with errors in mathematics classrooms. Structure and relevance of perceived error climate. *Contemporary Educational Psychology*, *38*(3), 196–210.

Tulis, M. (2013). Error management behavior in classrooms: Teachers‘ responses to student mistakes. *Teaching and Teacher Education*, *33*, 56–68.

Tulis, M., and Fulmer, S. M. (2013). Students’ motivational and emotional experiences and their relationship to persistence during academic challenge in mathematics and reading. *Learning and Individual Differences*, *27*, 35–46. https://doi.org/10.1016/j.lindif.2013.06.003.

Türling, J. M., Seifried, J., and Wuttke, E. (2012). Teachers’ knowledge about domain specific student errors. In E. Wuttke and J. Seifried (Eds.), *Research in vocational education: Vol. 1. Learning from Errors at School and at Work* (pp. 95–110). Opladen: Budrich.

Van de Vijver, F. J. R., van Hemert, D. A., and Poortinga, Y. H. (2008). Conceptual issues in multilevel models. In F. J. R. van de Vijver, D. A. van Hemert, and Y. H. Poortinga (Eds.), *Multilevel Analysis of Individuals and Cultures* (pp. 3–26). New York: Psychology Press.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., and Baggett, W. B. (2003). Why Do Only Some Events Cause Learning During Human Tutoring? *Cognition and Instruction*, *21*(3), 209–249. https://doi.org/10.1207/S1532690XCI2103_01.

Vosniadou, S., and Brewer, W. F. (1987). Theories of Knowledge Restructuring in Development. *Review of educational research*, *57*(1), 51–67.

Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, *54*, 427–450.

Wu, M. L., Adams, R. J., and Wilson, M. R. (1998). ConQuest: Multi-aspect test software [Computer Software]. Camberwell: Australian Council for Educational Research.

Zander, L., Kreutzmann, M., and Wolter, I. (2014). Constructive handling of mistakes in the classroom: The conjoint power of collaborative networks and self-efficacy beliefs. *Zeitschrift für Erziehungswissenschaften*, *17*, 205–223.

Affiliations

Julia Käfer [a], Susanne Kuger [b], Eckhard Klieme [a], Mareike Kunter [c]

[a] DIPF | Leibniz Institute for Research and Information in Education, Department of Educational Quality and Evaluation, Rostocker Straße 6, 60323 Frankfurt am Main, Germany

[b] German Youth Institute, Department of Social Monitoring and Methodology, Nockherstr. 2, 81541 München, Germany

[c] Goethe University Frankfurt, Faculty of Psychology and Sports Sciences, Department for Educational Psychology, Theodor-W.-Adorno-Platz 1, 60323 Frankfurt am Main, Germany

E-mail addresses:

Julia Käfer: julia.kaefer@dipf.de

Susanne Kuger: kuger@dji.de

Eckhard Klieme: klieme@dipf.de

Mareike Kunter: kunter@paed.psych.uni-frankfurt.de

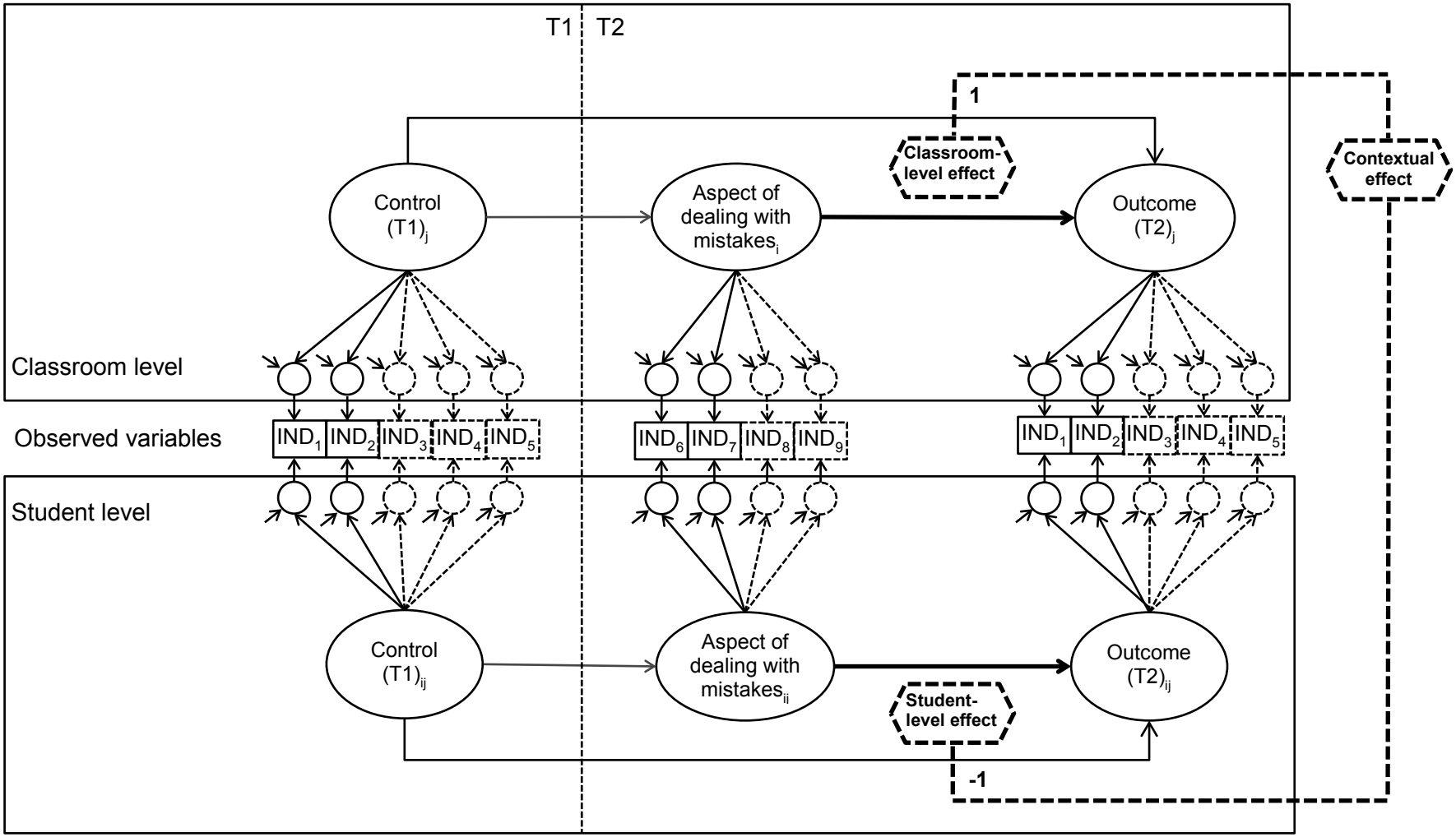

Fig. 1 Example of the ML-SEM models. Variables on the left-hand side are time 1 variables; variables on the right-hand side are time 2 variables. Control variables are either achievement at T1 or motivation at T1. Aspects of dealing with mistakes are teacher attitude toward mistakes, teacher response to student mistakes or students' perception of mistakes as useful for learning. Outcome variables are either achievement at T2 or motivation at T2. The subscripts ij at the student level indicate that the student-level variables may take on different values for each student i in each classroom j, whereas the subscript j at the classroom level indicate that classroom-level variables indicate the same value for each student in the same class but can vary between classrooms

Table 1: Item formulations, internal consistencies, means and standard deviations of the three aspects of dealing with mistakes used in the present study

Teacher attitude toward mistakes		
My English teacher is convinced that mistakes are useful because one can learn from them. My English teacher thinks that making mistakes is part of the language learning process. If my English teacher does not know something, he/she admits it openly.	Cronbach's α=0.73	*M*=2.78, *SD*=0.68
Teacher response to student mistakes		
If my written work goes wrong, my English teacher discusses the mistakes with me. If I use a wrong word, my English teacher explains to me why it is wrong. If I pronounce something wrong, my English teacher explains to me why this is wrong. If I make mistakes in English, I get the opportunity to correct myself or to start again.	Cronbach's α=0.76	*M*=2.75, *SD*=0.66
Students' perception of mistakes as useful for learning		
Making mistakes during English class helps me to do it better next time. I can learn more in English when I am allowed to make mistakes once in a while.	Cronbach's α=0.61	*M*=2.92, *SD*=0.67

The items were translated into English by the authors

Table 2: Correlations among the latent variables at student level and classroom level

	T1 achievement		T1 motivation		T2 achievement		T2 motivation		Teacher attitude		Teacher response		Perceived usefulness
Correlations at student level													
T1 achievement													
T1 motivation	*0.44*	(0.02)											
T2 achievement	*1.00*	(0.02)	*0.44*	(0.02)									
T2 motivation	*0.41*	(0.02)	*0.75*	(0.01)	*0.44*	(0.02)							
Teacher attitude	*0.13*	(0.02)	*0.19*	(0.02)	*0.18*	(0.02)	*0.24*	(0.02)					
Teacher response	*0.14*	(0.02)	*0.18*	(0.02)	*0.20*	(0.02)	*0.28*	(0.02)	*0.72*	(0.02)			
Perceived usefulness	*0.11*	(0.03)	*0.23*	(0.02)	*0.18*	(0.03)	*0.30*	(0.02)	*0.63*	(0.02)	*0.74*	(0.02)	
Correlations at classroom level													
T1 achievement													
T1 motivation	*0.74*	(0.04)											
T2 achievement	*1.00*	(0.01)	*0.74*	(0.04)									
T2 motivation	*0.68*	(0.04)	*0.92*	(0.03)	*0.70*	(0.04)							
Teacher attitude	*0.24*	(0.06)	*0.46*	(0.07)	*0.27*	(0.06)	*0.68*	(0.06)					
Teacher response	*0.12*	(0.06)	*0.35*	(0.07)	*0.14*	(0.06)	*0.56*	(0.06)	*0.92*	(0.03)			
Perceived usefulness	*0.49*	(0.07)	*0.54*	(0.08)	*0.53*	(0.07)	*0.74*	(0.07)	*0.93*	(0.06)	*0.91*	(0.05)	

N = 5266 students from 427 classes. T1 = Time 1; T2 = Time 2. Italicized parameters are statistically significant ($p < 0.05$).

Table 3: Influence of three aspects of dealing with mistakes on student achievement and motivation in English as a foreign language, after controlling for prior achievement and prior motivation, respectively. Results based on ML-SEM modelling. Standardized parameter estimates

	Achievement in EFL			Motivation in EFL		
	Model 1a	Model 1b	Model 1c	Model 2a	Model 2b	Model 2c
Student-level effects						
Aspects of dealing with mistakes						
Teacher attitude	*0.05* (0.02)			*0.10* (0.02)		
Teacher response		*0.06* (0.02)			*0.15* (0.02)	
Perceived usefulness			*0.08* (0.03)			*0.14* (0.02)
Control variables						
Prior knowledge	*1.00* (0.02)	*0.99* (0.00)	*0.99* (0.00)			
Prior motivation				*0.74* (0.01)	*0.73* (0.01)	*0.72* (0.01)
Classroom-level effects						
Aspects of dealing with mistakes						
Teacher attitude	0.02 (0.01)			*0.09* (0.03)		
Teacher response		0.00 (0.01)			0.04 (0.02)	
Perceived usefulness			0.02 (0.02)			*0.10* (0.03)
Control variables						
Prior knowledge	*1.00* (0.01)	*1.00* (0.00)	*0.97* (0.02)			
Prior motivation				*0.77* (0.05)	*0.82* (0.04)	*0.73* (0.07)

N = 5266 students from 427 classes. Loadings of latent factors are not given in the table. EFL = English as a foreign language. Italicized parameters are statistically significant ($p < 0.05$). The classroom-level effects of the three aspects of dealing with mistakes are controlled for the corresponding student-level effects

Table 4: Tests of measurement invariance via ML-CFA. Summary of goodness of fit statistics

Outcome	Model	χ^2	df	CFI	TLI	RMSEA	SRMR (W/B)		Description
Teacher attitude towards mistakes									
Achievement in EFL	M0AchAtt	225	22	0.966	0.935	0.042	0.026	0.042	No invariance
	M1AchAtt	247	23	0.962	0.931	0.043	0.026	0.042	Invariance over level for T2 achievement
	M2AchAtt	247	23	0.962	0.931	0.043	0.026	0.042	Invariance over level for T1 achievement
	M3AchAtt	263	24	0.960	0.930	0.043	0.027	0.087	Invariance over level for teacher attitude
	M4AchAtt	309	27	0.953	0.927	0.045	0.027	0.087	Invariance over level and over time for achievement
Motivation in EFL	M0MotAtt	1961	124	0.915	0.893	0.056	0.047	0.089	No invariance
	M1MotAtt	1914	128	0.917	0.899	0.055	0.046	0.092	Invariance over level for T2 motivation
	M2MotAtt	1918	128	0.917	0.899	0.055	0.047	0.093	Invariance over level for T1 motivation
	M3MotAtt	2014	126	0.912	0.891	0.057	0.047	0.117	Invariance over level for teacher attitude
	M4MotAtt	1967	138	0.915	0.904	0.053	0.046	0.111	Invariance over level and over time for motivation
Teacher response to student mistakes									
Achievement in EFL	M0AchRes	303	37	0.964	0.945	0.037	0.028	0.087	No invariance
	M1AchRes	321	38	0.961	0.943	0.038	0.028	0.087	Invariance over level for T2 achievement
	M2AchRes	318	38	0.962	0.944	0.037	0.028	0.087	Invariance over level for T1 achievement
	M3AchRes	334	40	0.960	0.944	0.037	0.027	0.098	Invariance over level for teacher response
	M4AchRes	375	43	0.955	0.941	0.038	0.028	0.098	Invariance over level and over time for achievement
Motivation in EFL	M0MotRes	1995	148	0.918	0.899	0.052	0.045	0.102	No invariance
	M1MotRes	1965	152	0.920	0.904	0.051	0.045	0.105	Invariance over level for T2 motivation
	M2MotRes	1971	152	0.919	0.904	0.051	0.045	0.109	Invariance over level for T1 motivation
	M3MotRes	2035	151	0.917	0.899	0.052	0.045	0.121	Invariance over level for teacher response
	M4MotRes	2010	163	0.918	0.909	0.049	0.044	0.115	Invariance over level and over time for motivation
Students' perception of mistakes as useful for learning									
Achievement in EFL	M0AchUse	215	15	0.949	0.898	0.050	0.030	0.069	No invariance
	M1AchUse	232	16	0.945	0.897	0.051	0.030	0.069	Invariance over level for T2 achievement
	M2AchUse	229	16	0.946	0.898	0.050	0.030	0.069	Invariance over level for T1 achievement
	M3AchUse	222	16	0.947	0.901	0.049	0.031	0.074	Invariance over level for perceived usefulness
	M4AchUse	258	19	0.939	0.904	0.049	0.032	0.075	Invariance over level and over time for achievement
Motivation in EFL	M0MotUse	1840	102	0.909	0.882	0.060	0.048	0.077	No invariance
	M1MotUse	1823	106	0.910	0.888	0.059	0.047	0.083	Invariance over level for T2 motivation
	M2MotUse	1806	106	0.911	0.889	0.059	0.047	0.079	Invariance over level for T1 motivation
	M3MotUse	1866	103	0.908	0.882	0.060	0.048	0.165	Invariance over level for perceived usefulness
	M4MotUse	1827	115	0.910	0.897	0.056	0.047	0.111	Invariance over level and over time for motivation

SRMR=Standardized Root Mean Square Residual, RMSEA=Root Mean Square Error of Approximation, TLI=Tucker-Lewis Index CFI=Comparative Fit Index. The robust Chi² test statistic were used as operationalized in Mplus by applying the MLR estimator (Muthén & Muthén, 1998-2013).W=within-group portion of the model; B=between-group portion of the model. EFL=English as a foreign language.