Unpacking domain-specific achievement motivation: the role of contextualising items for test-criterion correlations

Yvonne A. Michel\textsuperscript{a}, Ricarda Steinmayr\textsuperscript{b}, Anne C. Frenzel\textsuperscript{c} and Matthias Ziegler\textsuperscript{d}

\textsuperscript{a}Institute of Health and Society, University of Oslo, Oslo, Norway; \textsuperscript{b}Department of Psychology, Technical University Dortmund, Dortmund, Germany; \textsuperscript{c}Department of Psychology, Ludwig-Maximilians-University Munich, Munich, Germany; \textsuperscript{d}Institute for Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

ABSTRACT
Achievement motivation scores on the domain-specific level are better predictors of domain-matching scholastic performance than scores of general achievement motivation measures. Although there is research on domain-specific motivational measures, it is still unknown where this higher predictive power originates from. To address this, 715 students in secondary school answered questionnaires on general and domain-specific achievement motivation, domain-specific self-concept, and domain-specific self-esteem in two different studies. The first study was designed to disentangle the variance components in general and domain-specific achievement motivation in order to delineate hypotheses regarding potential drivers for the predictive power of domain-specific achievement motivation. The findings implied a strong role for a shared method factor. To explore the nature of this method factor, domain-specific self-concept/-esteem were focussed to establish discriminant validity evidence in a second study. The results indicate that the additional domain-specific variance can, in large parts, be explained by self-concept and self-esteem on domain-specific level.

The interplay between personal goals, self-beliefs, motives, and scholastic achievement is of growing interest to many researchers and practitioners, especially since the increasing recognition that motivational constructs contribute to the prediction of scholastic performance incrementally to intelligence and broader personality traits (Steinmayr & Spinath, 2007, 2009). Test scores from achievement motivation measures have been shown to have robust and replicable test-criterion correlations with criteria of scholastic performance (Richardson, Abraham, & Bond, 2012). Prior research suggested that scores from domain-specific achievement motivation measures outperform...
scores from general achievement motivation measures in terms of test-criterion correlations (Steinmayr & Spinath, 2009). Thus, through contextualising items by, for example, adding a phrase like ‘… in math’, achievement motivation scores correlate stronger with domain-specific performance. The present studies add to this literature by applying variance decomposition models to generally and domain-specifically phrased achievement motivation measures. These analyses allow to quantify trait variance and method variance due to contextualising, and to investigate the contribution of each variance source when predicting scholastic performance. Finally, we test whether the variance based on contextualising such achievement motivation items can be discriminated from other, achievement relevant constructs such as self-concept and self-esteem.

**Domain-specific achievement motivation and scholastic performance**

In their review, Murphy and Alexander (2000, p. 41) summarised that research on motivation develops ‘toward a more domain-specific viewpoint’. This, along with the increasing body of literature utilising domain-specific scales, was the rationale for this research project. As stated, a growing number of studies focus on domain-specific achievement motivation when predicting academic performance on domain level (Green, Martin, & Marsh, 2007; Wigfield, 1997; Wigfield, Guthrie, Tonks, & Perencevich, 2004). Domain-specific measures of achievement motivation play an important role in predicting interindividual differences in scholastic performance. Interestingly, the use of such contextualised items stands in contrast to traditional definitions for many of the constructs contextualisation is applied to. For example, Murray defined achievement motivation in a way that suggests that it acts the same way, regardless of the situation (Steinmayr & Spinath, 2009). However, research on the situational consistency of traits shows that even relatively stable traits, such as the Big Five, underlie situational variations (Deinzer et al., 1995). Moreover, adding contextualising phrases potentially changes the frame of reference (Schmit, Ryan, Stierwalt, & Powell, 1995; Shaffer & Postlethwaite, 2012) and thereby the answer process (Krosnick, 1999; Ziegler, 2011). Another relevant aspect is that contextualising changes the level of abstraction causing scale scores from contextualised items to be more symmetrical to many outcomes than scale scores from domain-general items (Kretzschmar, Spengler, Schubert, Steinmayr, & Ziegler, 2018). This is why, despite the inherent violation of underlying definitions, researchers before us have called for future studies which ‘… should compare the predictive power of the constructs assessed at the same level of generality and include further motivational constructs …’. A possible design should cover a complete crossover of the investigated motivational concepts and level of generality, i.e. need for achievement, ability self-concepts, values, and goals would be assessed with regard to motivation in general, at school, and for different subjects. Even though such a design would not necessarily assess each motivational construct in accordance with its theoretical foundation, it would allow for a more straightforward comparison of the predictive power of the concepts’ (Steinmayr & Spinath, 2009, p. 89). Wigfield and Cambria (2010), in a review paper, also summarised that the question of domain specificity of motivational measures is clear for some
constructs (e.g. self-efficacy) and less clear for others (e.g. task orientation) which in their opinion is mainly due to the lack of domain-specifically phrased measures for some of the constructs. Following the call by Steinmayr and Spinath (2009) and the other mentioned ideas, understanding the variance makeup of contextualised measures is an important step. Although the body of research on the predictive power of domain-specific motivational measures is growing, the only explanation for their relatively better test-criterion correlations so far is the level of symmetry between predictor and criteria (Brunswik, 1955; also see Bandura, 1997; Mischel, 1977). The key goal of the present research was to explore the change in score variance introduced by contextualising items by decomposing the variance into its different components, that is variance due to the trait being measured, variance due to contextualising (method variance), and measurement error.

**Different theoretical approaches to achievement motivation in the school context**

Among the variety of achievement-related motivational constructs, we concentrate on the need for achievement and goal-orientations (Murphy & Alexander, 2000). Clearly, the construct space in this field is much broader. However, we had to select a few constructs to pay tribute to the practical limitations of collecting data in schools. The constructs were therefore selected based on prior work showing their relevance for performance in different school subjects (Steinmayr & Spinath, 2007, 2009; Ziegler, Knogler, & Bühner, 2009; Ziegler, Schmidt-Atzert, Bühner, & Krumm, 2007). Accordingly, our results are limited to the selection made and future research needs to include a broader range of constructs.

**Need for achievement and scholastic performance**

Murray (1938) defined need for achievement as a drive to reach internal or external goals and called it a basic human need. McClelland, Atkinson, Clark, and Lowell (1953) differentiated the need for achievement into two achievement motives: Hope for success and fear of failure. They also characterised achievement situations by an emotional conflict in which hope for success and fear of failure have to be balanced. These motives are believed to arise before the actual achievement situation thereby determining the direction, intensity, and quality of achievement-related behaviour. While McClelland was highly skeptical of self-report measures (McClelland, Koestner, & Weinberger, 1989), over the years self-reports have become the standard means to capture both motives (e.g. Achievement Motivation Questionnaire by Nygård & Gjesme, 1973). McClelland et al. (1989) hypothesised that such self-reports of achievement motivation might be influenced by self-concept. Sparfeldt and Rost (2011) explored how well general versus domain-specific variants of hope for success and fear of failure predict grades in four subjects (mathematics, German, physics, and English). They found that domain-specific achievement motives showed stronger test-criterion correlations with their respective grades than general hope for success or fear of failure scores.
**Goal theories and scholastic performance**

Early achievement motivation theories assumed that capacity or ability could be seen based on mastery of a subject or based on performance comparisons (Nicholls, 1984). These thoughts were developed into the concepts of mastery and performance goals (Ames & Archer, 1987). While mastery orientation implies the desire of a person to gain competence, a performance-oriented person aims to outperform others or an external standard like grades. This dichotomous framework was further developed towards the $2 \times 2$ achievement goal framework in which a mastery-performance dimension is crossed with an approach-avoidance dimension (Elliot & McGregor, 2001). Approach goals motivate a person to approach situations in which competence can be shown, while avoidance goals make a person avoid such situations. The $2 \times 2$ framework thus postulates mastery-approach and mastery-avoidance, performance-approach and performance-avoidance goals. Its focus is on emotions experienced during or after achievement situations. Those can be assessed independently with the Achievement Goal Questionnaire (AGQ; Elliot & McGregor, 2001). The psychometric properties of the AGQ and the theoretical plausibility of the $2 \times 2$ framework have been shown in many studies (e.g. Elliot & Murayama, 2008). It has to be noted, though, that the mastery-avoidance scale often shows suboptimal psychometric properties, especially internal consistencies (Finney, Pieper, & Barron, 2004). Wirthwein, Sparfeldt, Pinquart, Wegerer, and Steinmayr, (2013) conducted a meta-analysis on the relations between goal-orientations and academic achievement. They found that matching specificity is a significant moderator, which again attests to the better test-criterion correlation of domain-specific scale scores.

**Variance decomposition as a way to analyse test-criterion correlations**

Despite the consistent empirical evidence showing the higher test-criterion correlations for scores from domain-specifically phrased measures compared to scores from generally phrased measures, explanations for this effect are scarce. We propose that there must be different variance sources within those scores, which show differential correlations with scholastic performance criteria. To identify such specific relations, it is helpful to first compare generally and domain-specifically phrased scales in terms of their variance components (Krumm, Hüffmeier, & Lievens, 2019). Items of these scales should be identical, except for the addition of ‘in math’, to make sure that they share the same content variance, which pertains to the actual trait measured (e.g. hope for success) and only differs in the methodological, context-specifying addition (i.e. generally versus in math).

When answering a contextualised item, test-takers are supposedly accessing a somewhat different construct compared to the generally phrased items. Current theory suggests that this construct is a domain-specific facet of the general construct. However, this explanation has not been formally tested. Nevertheless, modelling the different variance sources (i.e., trait variance and variance due to the added phrase) would allow two things: First, the actual amount of variance due to contextualising could be estimated. Second, the variance components could be correlated with
other measures, to shed light on their role in predicting criteria like scholastic performance.

One approach that seems suitable for this purpose is so called Correlated-Trait-Correlated-Method minus One model (CTC(M-1)) which was developed by Eid et al. (e.g. Eid et al., 2008). These models serve the purpose of decomposing scale variance into trait, method, and error variance into separate latent variables. For our particular case, this implies that both the general and the specific item, for example capturing hope for success, should be loaded by a latent trait factor hope for success, but the specific item would additionally be loaded by a corresponding latent domain method factor. Each item would have its individual latent error. Based on the loadings with the latent trait factor, a consistency coefficient can be estimated which reflects the amount of systematic variance within each measure due to the common trait in question (the variance due to the trait is divided by the sum of this variance and the variance due to the method). One important aspect of CTC(M-1) models is that they require choosing one method (e.g. generally or domain-specifically phrased) as the reference method (thus, the ‘minus 1’ in the model name). The variance due to this method is not specified as a latent variable in the models. One statistical advantage is that the models have fewer convergence problems (Eid et al., 2008). A CTC(M-1) model allows to estimate a domain-specific method factor, which is interpretable as the deviation of the domain-specific scale from the score which would be expected based on the reference method. CTC(M-1) models further allow estimating the amount of variance in each scale due to this method variable (method specificity: variance due to the method divided by the sum of this variance and the variance due to the trait). Moreover, it is possible to use the different variance components (trait versus method factor) as correlates of criteria. Thus, CTC(M-1) models inform us about the relative proportion of variance shared between the general and the domain-specific scale versions, in addition to non-shared variance (i.e. method variance due to contextualising the scales).

We conducted two studies, each using a systematic experimental item variation. We used items which only differed in their specificity (general versus in various specific domains) in combination with CTC(M-1) modelling. Whereas Study 1 focussed on the variance decomposition and the role each variance component has for test-criterion correlations, Study 2 explored possible correlates of the latent method factor to test its discriminant validity.

**Study 1**

Despite the increased recognition of domain-specific achievement motivation measures when predicting scholastic performance, it remains unclear what the specific role of variance due to contextualising items is with regard to test-criterion correlations. With this study we want to add to the existing literature by addressing four questions: First, are measure specific method factors necessary? Second, does the trait variance within general and domain-specific achievement measures show convergent validity? Third, what are the relative proportions each variance source has in general and domain-specifically phrased measures? Fourth, what is the role of each variance source for test-criterion correlations?
Method of Study 1

Sample and power

We collected data from 325 students from two different schools (174 girls), aged 14.32 on average (SD = 0.92). The majority (n = 174) attended German ‘Hauptschule’, a school which offers lower secondary education, n = 151 attended German ‘Gymnasium’, a school that offers higher secondary education in preparation for university. Students also differed in their grades with 194 students in 8th grade and 131 in the 9th grade.

Procedure and measures

Procedure

All data were collected via student self-report, during regular class hours on a voluntary basis after having received written consent from their parents. Most questionnaires needed to be adapted to get a general and a domain-specific version (see Appendix 2). Thus, a test battery was generated that included general versions of all questionnaires used as well as math specific versions of the same questionnaires. The latter was created by adding ‘in math’ or ‘in math class’ to general items. Descriptive statistics and internal consistencies for all measures can be found in Appendix 1. All internal consistency estimates were satisfactory with the exception of mastery-avoidance. As the modelling approach chosen uses latent variables and corrects for attenuation effects between latent variables, the score was used nevertheless. Moreover, there is empirical evidence suggesting that internal consistency estimates have little influence on test-criterion correlations focused here (McCrae, Kurtz, Yamagata, & Terracciano, 2011).

Measures

We used adapted versions of the approach versus avoidance by mastery versus performance motivation scales of Elliot and McGregor’s Achievement Goal Questionnaire (2001) and the revised German Achievement Motives Scale (AMS-R; Lang & Fries, 2006; cf. Nygård & Gjesme, 1973, see Appendix 2 for further details).

Scholastic performance. School grades in mathematics, German, and physics were reported by the students and used as criteria of scholastic performance. Grades range from 1, the best, to 6, the worst grade, with 5 and 6 indicating insufficient performance. We inverted the grades, in order to facilitate interpretation.

Statistical analyses

Models

R (R Core Team, 2016) and Mplus 8.1 (Muthén & Muthén, 1998–2012) were used for all analyses. Robust maximum-likelihood estimation (MLR) plus type two-level complex with school type as stratification and school class as cluster variables were used to analyse the structural equation models. These specifications help to deal with violations of the multivariate normal distribution as well as with the nested data structure due to the two school types and students being in different classes. Model fit was
judged on the basis of the $\chi^2$-test, the Standardised Root Mean Square Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA) as recommended by Hu and Bentler (1999), and on the Comparative Fit Index (CFI) as advised by Beauducel and Wittmann (2005). We used the following cut-off criteria for assuming good model fit: the SRMR should be smaller than 0.11 in combination with an RMSEA of less than or equal to 0.08 ($n > 250$), and a CFI value above 0.95.

The first CTC(M-1) model (Model 1) specified six correlated latent trait factors corresponding to the different theoretical approaches to achievement motivation: mastery-approach (MAP), mastery-avoidance (MAV), performance-approach (PAP), performance-avoidance (PAV), hope for success (HS), and fear of failure (FF) (Figure 1). Scores for each instrument and each level of globality (domain-specifically versus generally phrased) were also loaded by a method factor. In order to create a CTC(M-1) model, not all method factors were specified, though. As we considered the test family as a method, we decided to use the AMS as a reference method. Therefore, no method factor for the respective globally phrased scores was specified. Method factors for domain-specifically phrased scores were allowed to correlate. In Model 2, the same two factors were merged (Figure 2). The models were compared using the difference in CFI (Meade, Johnson, & Braddy, 2008). If Model 2 fitted better, this would mean that the additional variance due to adding ‘in math’ is not instrument-specific. In other words, it would not be facet variance of achievement motivation or goal-orientation but the very same variance in each of the scales which makes up the added variance due to the phrase ‘in math’.

Consistency and method-specificity were estimated following Nussbeck, Eid, Geiser, Courvoisier, and Lischetzke (2009). Accordingly, consistency can be understood as the
proportion of trait variance relative to trait and method factor variance. Method specificity is the proportion of method factor variance relative to trait and method factor variance. Confidence intervals for all paths were estimated (95% intervals) in order to compare the loadings for generally and domain-specifically phrased scales.

In a second analytical step, different school grades (math, physics, German) were integrated into the preferred model (Models 2a–c) and regressed on the latent trait and method factor(s). For the math grade, this model informs us about the specific role each variance source has for the test-criterion correlation. As for the physics and German grade, those models will show whether this specific role is related to the domain itself or can be generalised across domains of scholastic performance.

Results of Study 1

**CTC(M-1): model fit, consistency, method specificity, and convergent validity**

Model 1 with three instrument specific method factors yielded an acceptable fit ($\chi^2_{[100]} = 324.45, p < .001$, RMSEA = 0.083, CFI = 0.937, SRMR = 0.034). Model 2 in which the method factors were merged for the domain-specifically phrased scores had a very similar fit ($\chi^2_{[101]} = 325.81, p < .001$, RMSEA = 0.083, CFI = 0.937, SRMR = 0.034). Thus, assuming the same latent source behind the different instruments did not deteriorate model fit. Consistencies and specificities from this model can be found in Table 1. It has to be noted here that the specificities for all globally phrased scores were zero. Thus, the method variable had no influence and was therefore not modelled in subsequent analyses.
| Method factor | HS (g) 0.84*** (0.80–0.88) | M_{AP} (g) 0.84*** (0.83–0.85) | M_{AV} (g) 0.64*** (0.59–0.69) | P_{AP} (g) 0.93*** (0.88–0.98) | P_{AV} (g) 0.84*** (0.81–0.87) | FF (g) 0.86*** (0.81–0.91) | Method factor loading | Consistency | Method-specificity |
|---------------|---------------------------|-------------------------------|---------------------------------|-------------------------------|-------------------------------|--------------------------|-------------------|---------------|------------------|
| HS (d) 0.68*** (0.64–0.72) | M_{AP} (d) 0.66*** (0.62–0.70) | M_{AV} (d) 0.58*** (0.50–0.66) | P_{AP} (d) 0.92*** (0.88–0.94) | P_{AV} (d) 0.83*** (0.80–0.86) | FF (s) 0.79*** (0.76–0.82) | FF (s) –0.15*** | 10.00           | 0.56***       | 0.60             | 0.40             |
|               |                           |                               |                                 |                               |                               |                          |                   |               |                 |                 |
|               |                           |                               |                                 |                               |                               |                          |                   |               | 0.59***        | 0.56             | 0.44             |
|               |                           |                               |                                 |                               |                               |                          |                   |               | 0.26***        | 0.83             | 0.17             |
|               |                           |                               |                                 |                               |                               |                          |                   |               | 0.40***        | 0.81             | 0.19             |
|               |                           |                               |                                 |                               |                               |                          |                   |               | 0.83***        | 0.87             | 0.13             |
|               |                           |                               |                                 |                               |                               |                          |                   |               | 0.87***        | 0.90             | 0.10             |

Note. HS: hope for success; FF: fear of failure; MAP: mastery-approach; MAV: mastery-avoidance; PAP: performance-approach; PAV: performance-avoidance; (g): general; (d): domain. **p ≤ .001. 95% confidence intervals can be found in parentheses below each estimate.
Test-criterion correlations: role of trait and method factors

Table 2 shows the results of regressing grades onto the latent trait variables and the method factor. It can be seen that fear of failure was related to the math grade, hope for success to the German grade (see Steinmayr & Spinath, 2009 for similar results), and the mastery approach to the physics grade. While the size of the regression weights did not differ much across domains, the standard errors and thus, the significances varied. Consequently, any conclusions regarding the importance of specific traits for specific domains should be postponed until the complete patterns have been replicated. An interesting aspect here could be the difference between math and physics on the one side and German, as the native language, on the other side.

Of specific interest to the current study was the strong regression weight of the method factor for the math grade. The variance component due to the phrase ‘in math’ was strongly related to math performance but not the other grades.

Discussion of Study 1

In Study 1, we aimed at decomposing the variance of scores from generally and domain-specifically phrased achievement motivation measures. Using CTC(M-1) models, the variance was decomposed into three parts. The first part reflects interindividual trait differences. The second part contains the variance due to contextualising items. Finally, unsystematic measurement error was also modelled. The results show that the variance added by including the phrase ‘in math’ is not instrument-specific (research question 1). The model in which only one latent variable was used fitted equally well. Moreover, the influence of the method on generally phrased items was negligible. Second, the findings attest to strong trait loadings and convincing convergent validity between the corresponding domain and generally phrased scores (research question 2). However, the models in which grades were predicted imply that the incremental validity of domain-specifically phrased scale scores comes from the variance introduced by ‘in math’ (research question 3). The findings further show that adding this phrase had more influence on positively connoted traits (hope for success and mastery approach) than negatively connoted traits. The results also replicate the stronger test-criterion correlations with the math grade for domain-specifically phrased scales on the manifest level. Within the CTC(M-1) models, it could be seen that especially the
variance captured in the method factor is related to variance in math grades (research question 4). This specific relation does not generalise across grades.

Study 2 will follow up on these findings and test how the method variance due to item contextualisation relates to other constructs.

**Limitations**

Despite the relative breadth in constructs, we operationalised each construct with one measure only. This was done to keep the total number of items at a minimum. However, the results show that consistency and specificity varied. Thus, different measures from the ones used here might also yield different relative variance proportions. Second, we only operationalised domain-specificity with regard to math. Whether the findings generalise to contextualising items for other school subjects remains an open question, which we set out to tackle in Study 2. Finally, the sample consisted of rather young students from heterogeneous school types. While we tried to control for the latter statistically, both aspects might have led to variance inflations which would lead to overestimations. Consequently, Study 2 utilised a more homogeneous sample with regard to age and school type.

To sum up, Study 1 revealed that domain-specifically phrased instruments include substantial amounts of variance which can be interpreted as results of contextualising items by including the phrase ‘in math’. It could be shown that this variance component is not instrument-specific. However, it was related to scholastic performance. The nature of this variance, modelled as a method factor here, remains unclear, even though positively connoted traits seem to be more affected by it.

**Study 2**

The aims of Study 2 were to replicate the findings of Study 1 using a sample of older and with regard to school type more homogeneous students. At the same time, items were contextualised not only for math but also for German. Moreover, based on theoretical grounds, we explored possible correlates of the method factor.

**Possible correlates of the method factor**

Study 1 revealed that the variance captured in the method factor is related to grades in the respective school subject, at least for math. Thus, when it comes to gauging the nature of this method factor, two alternative hypotheses seem viable. First, it could be assumed that the deviation from the generally phrased scales reflects a person’s standing on a lower order facet of the trait. Alternatively, it could be assumed that by contextualising items, additional constructs are being tapped. This raises the question as to which constructs might be tapped when adding phrases such as ‘in math’.

We selected potential additional constructs based on two aspects. First, as McClelland et al. (1989) already proposed, self-reported achievement motivation might be nothing but self-concept. With regard to domain-specific achievement motivation,
the additional method variance might tap into domain-specific self-concepts. Research on self-concept shows that it is organised in a domain-specific way, relevant to the prediction of scholastic performance, and related to achievement motivation (Helmke & van Aken, 1995; Jansen, Schroeders, & Lüdtke, 2014, Marsh & O’Mara, 2008; Suárez-Alvarez, Fernández-Alonso, & Muniz, 2014). An interesting observation is that most self-concept scales achieve domain-specificity by contextualising items in the same way as it is done for achievement motivation scales, i.e. by adding phrases such as ‘in math (class)’ (e.g. the widely used Academic Self-Description Questionnaire II by Marsh, 1990b).

Another similarity between self-concept scales and the method factor specified here is that both overlap with differences in grades. Earlier findings suggest that not only self-concept, but also self-esteem is related to grades, indicating that self-concept and self-esteem are distinguishable constructs (Jansen, Scherer, & Schroeders, 2015; Scherer, 2013). For investigating whether contextualising items by adding ‘in math’ introduces variance from additional constructs, both domain-specific self-concept and self-esteem are promising candidates. As several reviews underline the difficulties to distinguish self-concept and self-esteem, both were included here (Byrne, 1996; Valentine, DuBois, & Cooper, 2004). Finally, both constructs are positively connoted which fits the results of Study 1.

**Analytical strategy**

Ziegler and Hagemann (2015) highlighted the importance of testing the discriminant validity of test scores by including target and supposedly discriminant constructs within the same structural model. Ziegler and Bäckström (2016) also suggested following this approach when looking at test-criterion correlations of facets. This way, specific relations controlled for unwanted construct overlap can be identified. Based on the explanations above and these suggestions, it seems promising to test whether the variance added through contextualising achievement motivation scales can be discriminated from self-concept and self-esteem when looking at test-criterion correlations. Such a head-to-head comparison should leave sufficient specific variance to maintain a correlation between the method factor identified in Study 1 and scholastic performance criteria, if the method factor reflects something different from self-concept and self-esteem.

**Prior research on the interplay of motivational variables as predictors of specific school grades**

Research looking at domain-specific measures of motivational constructs has a long and rich tradition. While a lot of work directly focuses on test-criterion correlations (e.g. Bong, 2002a, 2002b; Lee, Lee, & Bong, 2014; Pajares & Miller, 1994, 1995; Steinmayr & Spinath, 2007, 2009) other research has focussed on more psychometric aspects. For example, temporal stability (Bong, 2005) or domain-specificity (Bong, 2001, 2002a, 2002b; Green et al., 2007) have been scrutinised. Those studies show that there is temporal stability and strong convergent validity evidence for different
domain-specific scales. Other studies have looked at common antecedents of domain-specific motivational constructs (Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006) and identified domain-specific motivational styles (Galloway, Leo, Rogers, & Armstrong, 1996) or epistemological beliefs (Buehl & Alexander, 2005). There is also research investigating gender-specific differences in test-criterion correlations (Spinath, Eckert, & Steinmayr, 2014; Spinath, Freudenthaler, & Neubauer, 2010). Valentine et al. (2004) reviewed results on the relation between self-beliefs, including self-concept, self-esteem, and self-efficacy, and achievement on general and domain-specific levels. There was a small but significant regression coefficient ($\beta = 0.08$) between self-beliefs and academic achievement on general level when controlling for prior achievement. The relation grew to a medium effect size for domain-specific measures. Seaton, Parker, Marsh, Craven, and Yeung (2014) investigated the interplay of math specific self-concept and math specific achievement motivation as predictors of math performance. Their findings were that math specific self-concept and math specific achievement motivation were moderately correlated as well as that math specific achievement motivation was not predictive of math performance over and above math specific self-concept. This is of relevance for Study 2. A replication of Seaton et al.'s findings would further support the hypothesis that adding 'in math' also adds another construct to the measures of achievement motivation, rather than capturing a facet level. However, despite all of this important and fruitful work, a systematic cross-over, comparing domain-specifically phrased with general measures using modern variance decomposition methods to take a closer look at the source of test-criterion correlations, is missing.

**Aims of Study 2**

The first aim of Study 2 was to replicate results from Study 1 and overcome some of its limitations. The second aim of Study 2 was to put domain-specific achievement motivation, self-concept, and self-esteem in a head-to-head comparison regarding test-criterion correlations. To this end, we again applied a CTC(M-1) approach to test whether the method variance due to adding phrases like 'in math' would still predict scholastic performance after controlling for self-concept and self-esteem.

**Method of Study 2**

**Sample and power**

The sample consisted of 390 students attending a German secondary school type that prepares for university (Gymnasium) (231 girls, average age 17 years [$SD = 1.06$]). The sample was recruited at two schools in two midsized towns, each testing two subsequent cohorts. Students participated voluntarily and their parents signed consent forms.

Again, a-priori power analyses considered the CTC(M-1) models and the regression. For the CTC(M-1) model a sample size of $N = 402$ was estimated to have sufficient power for a model with 5 degrees of freedom. The current sample size is just below the maximum of $N = 402$ needed.
**Measures**

Descriptive statistics and internal consistencies can be found in Appendix 4.

**Achievement motivation**

A German version (Dahme, Jungnickel, & Rathje, 1993) of the Achievement Motives Scale (AMS) by Nygård and Gjesme (1973) was applied to measure general and specific achievement motivation via self-report. See Appendix 3 for more information about the scale and its adaptation.

**Academic self-concept**

Academic self-concept was assessed by the Scales for the Assessment of Academic Self-Concept (SESSKO; Schöne, Dickhäuser, Spinath, & Stiensmeier-Pelster, 2003) in a domain-specific version (Steinmayr & Spinath, 2007, 2008, 2009, see Appendix 3 for further information).

**Scholastic performance**

As in Study 1, self-reported grades (from 1 to 6) in math and German were used as criteria for scholastic performance. As before, grades were inverted to facilitate interpretation.

**Statistical analyses**

As in Study 1, a robust maximum-likelihood estimator (MLR) was used. Missing values were dealt with by using the full information maximum likelihood method (FIML). The CTC(M-1) model tested contained two correlated latent trait variables, i.e. fear of failure and hope for success. The general scale version was chosen as the reference method. Thus, two method factors were specified, one representing deviations from the reference method when contextualising items for math and one representing deviations when contextualising items for German. Based on this model we estimated consistency and method-specificity. In the next step, we included either the math or the German grade and used it in a hierarchical latent regression. In the first block, we entered both latent traits and the domain-specific latent method factor as predictor variables. In a second block, we entered the respective domain-specific self-esteem and self-concept scores as additional predictor variables. Model fit was evaluated as in Study 1.

**Results of Study 2**

Table 3 shows the correlations between all variables. The main pattern for hope for success was that the higher a person scored on this subscale, the better the grade in this domain. For fear of failure, this pattern was reversed. As before, the correlations were higher when grade and scale score matched with regard to the domain. With regard to self-esteem, the correlations between different domain scores were comparable to prior research (Bong, 2002a, 2002b; Bong & Hocevar, 2002) as were the correlations to criteria (Pajares & Miller, 1994, 1995).
The specified CTC(M-1) model fit the data well: \( \chi^2(5) = 8.82, p < .117 \), RMSEA = 0.044, CFI = 0.992, SRMR = 0.028. Consistency was generally high and method-specificity low (Table 4). Nevertheless, there were substantial loadings from the latent method factor for each domain. Table 5 shows standardised regression weights for each block in both hierarchical latent regressions. It can be seen that both trait variables, as well as method factors, were related to the respective grade. All in all, the amount of explained variance for math grades was larger, mainly due to the regression weight of the method factor.
When adding domain-specific self-concept and self-esteem scores, the amount of explained variance for the grades increased strongly for each domain. Both domain-specific self-concept and self-esteem had substantial and significant regression weights in all domains. In addition, with self-concept and self-esteem scores in the models, the relation between the grades and the respective method factors was strongly reduced and no longer significant.

**Discussion of Study 2**

Study 2 was conducted in order to replicate findings from Study 1 using a more homogeneous sample with regard to age and school type as well as an additional contextualisation (‘in German’). The CTC(M-1) model again yielded convergent validity evidence. The amounts of consistency were roughly the same. The method factor variance again had substantial relations with grades. Most importantly, this relation was strongly reduced when domain-specific self-concept and self-esteem scores were controlled for.

**Correlates of the method factors**

The current results show that the role of each domain-specific method factor for the test-criterion correlation strongly deteriorates when controlling for domain-specific self-concept and self-esteem. Moreover, the scores for those constructs not only have strong regression weights, their inclusion substantially improves the regression model. Based on the procedural suggestions by Ziegler and Bäckström (2016) as well as Siegling, Petrides, and Martskvishvili (2015), this implies that the variable shown to be inferior in a head-to-head comparison might be redundant. Study 2 also confirms earlier findings (Seaton et al., 2014).

At least two alternative explanations for the findings and thus, arguments against the idea that domain-specific achievement motivation items capture domain-specific self-concept and self-esteem could be brought forward. First of all, it could be argued that it is actually the other way around, i.e. domain-specific measures of self-concept and self-esteem actually capture achievement motivation. Considering the well-replicated findings regarding the structure of self-concept this seems unlikely, though. The idea is further contradicted by the fact that domain-specific self-concept and self-esteem both retained most of their test-criterion correlations in the regression analyses while the method factor, allegedly domain-specific achievement motivation, did not. Second, it could be argued that the present findings simply show that domain-specific achievement motivation, self-concept, and self-esteem items share method variance due to the fact that all measures use the same phrase such as ‘in math’. However, findings from Pomerance and Converse (2014) can be used to refute such an alternative explanation. In contrast to our findings in the domain of motivation, they found that school-specific extraversion and conscientiousness scores kept their test-criterion correlations when domain-specific self-concept scores were added as additional predictors of scholastic performance, leadership, and health. The fact that the contextualised personality trait scores kept their test-criterion correlations makes the assumption...
unlikely that method variance due to adding ‘in math’ causes an overlap which decreases incremental validity.

**Limitations**

We only used two different operationalisations of achievement motivation, as the data were collected with several research questions in mind, thereby requiring a compromise regarding the total number of instruments. As a consequence, we could not replicate the finding that the method factor was not instrument-specific. Thus, future research needs to aim at replicating this. Moreover, self-concept and especially self-esteem were assessed with very few items. Thus, their role might be underestimated. However, recent research on test-criterion correlations of short scales supports the test-criterion related validity of even very short scales (Thalmayer, Saucier, & Eigenhuis, 2011; Ziegler, Poropat, & Mell, 2014).

One potential limitation of both studies not discussed so far is the use of self-reported grades instead of achievement test results. Such grades are potentially less objective. However, several studies showed that self-reported grades provide a valid representation of students’ actual performance (Dickhäuser & Plenter, 2005; Kuncel, Credé, & Thomas, 2005).

**General discussion**

The current research project aimed at adding to the literature on domain-specific achievement motivation and goal-orientation scales. Prior research had shown that contextualising such instruments yields improved test-criterion correlations. Before our studies, there was little empirical evidence supporting the claim that by contextualising items, specific facets of those traits are being tapped. In order to overcome this research gap, we utilised CTC(M-1) models in order to decompose scale variance and correlate the different variance components with grades. The results show that adding phrases such as ‘in math’ adds systematic variance which is related to grades and that this relation strongly deteriorates when controlling for self-esteem and self-concept in the respective domain. Moreover, the results imply that the additional variance is not instrument- or construct specific. Together these findings cast doubts on the interpretation of the additional variance component as reflecting domain-specific trait facets of achievement motivation.

Thus, the open question is, what exactly is being added by contextualising achievement motivation and goal-orientation items? Using models of the item response process we generate a hypothesis for future research: After a mental representation of the item content is built, it is compared to information retrieved from memory and a general judgment is made and mapped onto the response scale (Krosnick, 1999; Ziegler, 2011). According to the present findings, we hypothesise that the mental representation of a domain-specific achievement motivation or goal-orientation item including a phrase like ‘in math’ might evoke the retrieval of information reflecting a general achievement motive or general goal-orientation and, in addition, information on domain-specific self-concept and self-esteem. Taking the general level as a starting
point, the judgment could then be fine-tuned by domain-specific self-concept and self-esteem. This hypothesis would be in line with the idea of a CTC(M-1) model which specifies the added variance as the deviation from the scores expected based on the general trait. There is also evidence from other studies supporting this hypothesis. Duda and Nicholls (1992) showed that while beliefs regarding the necessity of interest, cooperation, and effort as underlying fundamentals of achievement motivation are comparable for sports and academic motivation, ability beliefs differ by domain. This would reflect the hypothesised answer process in which general motivation beliefs are coupled with domain-specific self-concept and self-efficacy.

Of course, our hypothesis does not necessarily mean that there is no such thing as a domain-specific achievement motivation or goal-orientation facet. In fact, one could argue that the results might be different for constructs not operationalised here. Still, the current research implies that using phrases such as ‘in math’ contextualises items, and that this contextualisation might not reflect differences on facet level for the constructs operationalised here. Based on our results, it seems more likely that additional variance due to different constructs like domain-specific self-esteem and self-concept is tapped. The fact that these findings might be confined to the constructs of achievement motivation and goal-orientation is highlighted by prior research on contextualising personality items which did not yield comparable results.

It has to be kept in mind that we only present limited empirical support for these assumptions. Thus, replications and thorough tests of our hypothesis are needed. Until our findings are replicated, it might be advisable to include motivational as well as self-concept and self-efficacy measures into a test battery to fully exploit the potentially predictive variance and also to find specific effects.

**Conclusion**

Prior research showed that domain-specifically phrased achievement motive scale scores yield better test-criterion correlations when predicting scholastic performance. Our studies built on this research and aimed to contribute towards a clearer picture of the specific contributions to test-criterion correlations different variance components in those test scores might have. Our results suggest that the higher test-criterion correlations of the domain-specific instruments could be related to additional, domain-specific variance captured. However, our findings also imply that this variance strongly overlaps with interindividual differences in domain-specific self-concept and self-esteem. If these findings were replicated, the idea of domain-specific achievement motivation facets would have to be seriously scrutinised.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

We acknowledge support by the Open Access Publication Fund of Humboldt Universität zu Berlin.
References

Ames, C., & Archer, J. (1987). Mothers’ beliefs about the role of ability and effort in school learning. *Journal of Educational Psychology, 79*, 409–414. doi:10.1037/0022-0663.79.4.409

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman New York.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(1), 41–75. doi:10.1207/s15328007sem1201_3

Bong, M. (2001). Between-and-within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology, 93*(1), 23–34. doi:10.1037/0022-0663.93.1.23

Bong, M. (2002a). Stability and structure of self-efficacy, task-value, and achievement goals and consistency of their relations across specific and general academic contexts and across the school year. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 1–5, 2002.

Bong, M. (2002b). Predictive utility of subject-, task-, and problem-specific self-efficacy judgments for immediate and delayed academic performances. *The Journal of Experimental Education, 70*, 133–162. doi:10.1080/00220970209599503

Bong, M. (2005). Within-grade changes in Korean girls’ motivation and perceptions of the learning environment across domains and achievement levels. *Journal of Educational Psychology, 97*, 656–672. doi:10.1037/0022-0663.97.4.656

Bong, M., & Hocevar, D. (2002). Measuring self-efficacy: Multitrait-multimethod comparison of scaling procedures. *Applied Measurement in Education, 15*, 143–171. doi:10.1207/S15324818AME1502_02

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217. doi:10.1037/h0047470

Buehl, M. M., & Alexander, P. A. (2005). Motivation and performance differences in students’ domain-specific epistemological belief profiles. *American Educational Research Journal, 42*, 697–726. doi:10.3102/0028312042004697

Byrne, B. M. (1996). Academic self-concept: Its structure, measurement, and relation to academic achievement. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 287–316). Oxford, England: John Wiley & Sons.

Dahme, G., Jungnickel, D., & Rathje, H. (1993). Güttigenschaften der Achievement Motives Scale (AMS) von Gjesme und Nygard (1970) in der deutschen Übersetzung von Göttert und Kuhl—Vergleich der Kennwerte norwegischer und deutscher Stichproben [Psychometric properties of the Achievement Motives Scale (AMS) by Gjesme and Nygard (1970) in the German Translation by Göttert and Kuhl]. *Diagnostica, 39*, 257–270.

Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F., & Neubauer, A. (1995). Situational effects in trait assessment: The FPI, NEO-FFI, and EPI questionnaires. *European Journal of Personality, 9*(1), 1–23. doi:10.1002/per.2410090102

Dickhäuser, O., & Plenter, I. (2005). Letztes Halbjahr stand ich zwei [Last semester I had a B]. *Zeitschrift Für Pädagogische Psychologie, 19*, 219–224. doi:10.1024/1010-0652.19.4.219

Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept [The Academic Self-concept scales]. *Zeitschrift Für Differentielle Und Diagnostische Psychologie, 23*, 393–405. doi:10.1024/0170-1789.23.4.393

Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology, 84*, 290–299.

Eid, M., Nussbeck, F., Geiser, C., Cole, D., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods, 13*, 230–253.

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501–519. doi:10.1037/0022-3514.80.3.501

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*, 613–628. doi:10.1037/0022-0663.100.3.613
Finney, S. J., Pieper, S. L., & Barron, K. E. (2004). Examining the psychometric properties of the Achievement Goal Questionnaire in a general academic context. *Educational and Psychological Measurement, 64*, 365–382. doi:10.1177/0013164403258465

Galloway, D., Leo, E. L., Rogers, C., & Armstrong, D. (1996). Maladaptive motivational style: The role of domain specific task demand in English and mathematics. *British Journal of Educational Psychology, 66*, 197–207. doi:10.1111/j.2044-8279.1996.tb01189.x

Green, J., Martin, A. J., & Marsh, H. W. (2007). Motivation and engagement in English, mathematics and science high school subjects: Towards an understanding of multidimensional domain-specificity. *Learning and Individual Differences, 17*, 269–279. doi:10.1016/j.lindif.2006.12.003

Helmke, A., & van Aken, M. A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology, 87*, 624–637. doi:10.1037/0022-0663.87.4.624

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

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students’ self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology, 41*, 13–24. doi:10.1016/j.cedpsy.2014.11.002

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*, 11–21. doi:10.1016/j.lindif.2013.12.003

Kretzschmar, A., Spengler, M., Schubert, A.-L., Steinmayr, R., & Ziegler, M. (2018). The relation of personality and intelligence—What can the Brunswik Symmetry Principle tell us? *Journal of Intelligence, 6*(3), 30. doi:10.3390/jintelligence6030030

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537–567. doi:10.1146/annurev.psych.50.1.537

Krumm, S., Höffmeier, J., & Lievens, F. (2019). Experimental test validation: Examining the path from test elements to test performance. *European Journal of Psychological Assessment, 35*, 225–232.

Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality, 43*, 489–493.

Kuncel, N., Credé, M., & Thomas, L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*(1), 63–82. doi:10.3102/00346543075001063

Lang, J. W. B., & Fries, S. (2006). A revised 10-item version of the Achievement Motives Scale: Psychometric properties in German-speaking samples; Eine revidierte 10-Item-Version der Achievement Motives Scale (Skala zu Leistungsmotiven). *European Journal of Psychological Assessment, 22*, 216–224. doi:10.1027/1015-5759.22.3.216

Lee, W., Lee, M.-J., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology, 39*, 86–99. doi:10.1016/j.cedpsych.2014.02.002

Marsh, H. W. (1990b). *SDQ II Manual: Self-Description Questionnaire-II*. Macarthur: University of Western Sydney.

Marsh, H. W., & O’Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*, 542–552. doi:10.1177/0146167207312313

McClelland, D., Atkinson, J., Clark, R., & Lowell, E. (1953). *The achieving motive*. New York: Appleton-Century Crofts.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*, 690–702. doi:10.1037/0033-295X.96.4.690

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*(1), 28–50. doi:10.1177/1088868310366253
Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*, 568–592. doi:10.1037/0021-9010.93.3.568

Mischel, W. (1977). On the future of personality measurement. *American Psychologist, 32*, 246–254. doi:10.1037/0003-066X.32.4.246

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25*(1), 3–53. doi:10.1016/ceps.1999.1019

Muthén, L. K., & Muthén, B. (1998–2012). *MPlus user’s guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328–346. doi:10.1037/0033-295X.91.3.328

Nussbeck, F. W., Eid, M., Geiser, C., Courvoisier, D. S., & Lischetzke, T. (2009). A CTC(M − 1) model for different types of raters. *Methodology, 5*, 88–98. doi:10.1027/1614-2241.5.3.88

Nygaard, R., & Gjesme, T. (1973). Assessment of achievement motives: Comments and suggestions. *Scandinavian Journal of Educational Research, 17*(1), 39–46. doi:10.1080/0031383730170104

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*, 193–203. doi:10.1037/0022-0636.86.2.193

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology, 42*, 190–198. doi:10.1037/0022-0167.42.2.190

Pomerance, M. H., & Converse, P. D. (2014). Investigating context specificity, self-schema characteristics, and personality test validity. *Personality and Individual Differences, 58*, 54–59. doi:10.1016/j.paid.2013.10.005

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students’ academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353–387. doi:10.1037/a0026838

Rost, J., Carstensen, C. H., & von Davier, M. (1999). Are the Big Five Rasch scalable? A reanalysis of the NEO-FFI norm data. *Diagnostica, 45*, 119–127.

Scherer, R. (2013). Further evidence on the structural relationship between academic self-concept and self-efficacy: On the effects of domain-specificity. *Learning and Individual Differences, 28*, 9–19. doi:10.1016/j.lindif.2013.09.008

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607–623. doi:10.1037/0021-9010.80.5.607

Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2003). Das Fähigkeitsselbstkonzept und seine Erfassung [Ability self-concept and its measurement]. *Diagnostik Von Motivation Und Selbstkonzept, 1*, 3–14.

Seaton, M., Parker, P., Marsh, H. W., Craven, R. G., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: Juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology, 34*(1), 49–72.

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*, 445–494.

Siegling, A. B., Petrides, K. V., & Martskivshvili, K. (2015). An examination of a new psychometric method for optimizing multi-faceted assessment instruments in the context of trait emotional intelligence. *European Journal of Personality, 29*(1), 42–54. doi:10.1002/per.1976

Sparfeldt, J. R., & Rost, D. H. (2011). Content-specific achievement motives. *Personality and Individual Differences, 50*, 496–501. doi:10.1016/j.paid.2010.11.016
## Appendix

### Appendix 1. Descriptive statistics, internal consistencies, and correlational patterns of general and domain-specific achievement motivation and grades.

|                | General |          |          |          |          |          | Domain |          |          |          |          |          |          |          |          |          |          |          |          |
|----------------|---------|----------|----------|----------|----------|----------|--------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
|                |         | $M$      | $SD$     | $R$      |          |          | $R$    |          |          |          |          |          |          |          |          |          |          |          |
| **General**    |         |          |          |          |          |          |        |          |          |          |          |          |          |          |          |          |          |          |
| $P_{Ap}$       | 1.7     | .61      |          |          |          |          |        |          |          |          |          |          |          |          |          |          |          |          |
| $P_{Av}$       | 1.7     | .67      | 0.79***  |          |          |          |        |          |          |          |          |          |          |          |          |          |          |          |
| $M_{Ap}$       | 2.0     | .52      | 0.66***  | 0.77***  |          |          |        |          |          |          |          |          |          |          |          |          |          |          |
| $M_{Av}$       | 1.5     | .58      | 0.38***  | 0.21***  | 0.66***  |          |        |          |          |          |          |          |          |          |          |          |          |          |
| $HS$           | 2.0     | .52      | 0.23***  | 0.19***  | 0.33***  | 0.43***  | 0.67*** |          |          |          |          |          |          |          |          |          |          |          |
| $FF$           | 1.2     | .62      | 0.26***  | 0.11*    | 0.46***  | 0.14*    | 0.67*** | 0.71***  |          |          |          |          |          |          |          |          |          |          |
| **Domain**     |         |          |          |          |          |          |        |          |          |          |          |          |          |          |          |          |          |          |          |
| $P_{Ap}$       | 1.6     | .69      | 0.57***  | 0.37***  | 0.21***  | 0.25***  | 0.04   | 0.85***  |          |          |          |          |          |          |          |          |          |          |          |
| $P_{Av}$       | 1.6     | .75      | 0.66***  | 0.72***  | 0.25***  | 0.21**   | 0.11   | 0.20*    | 0.74***  | 0.84***  |          |          |          |          |          |          |          |          |          |
| $M_{Ap}$       | 1.9     | .68      | 0.36***  | 0.22***  | 0.60***  | 0.28***  | 0.40*** | 0.11    | 0.55***  | 0.36***  | 0.84***  |          |          |          |          |          |          |          |          |
| $M_{Av}$       | 1.6     | .66      | 0.18***  | 0.22**   | 0.25***  | 0.37***  | 0.13*  | 0.20***  | 0.30***  | 0.34***  | 0.37***  | 0.57***  |          |          |          |          |          |          |          |
| $HS$           | 1.9     | .66      | 0.31***  | 0.21***  | 0.40***  | 0.12*    | 0.58*** | 0.06    | 0.47***  | 0.34***  | 0.66***  | 0.27***  | 0.80***  |          |          |          |          |          |          |
| $FF$           | 1.1     | .75      | 0.05     | 0.15     | 0.25***  | 0.14*    | 0.03   | 0.74***  | −0.02    | 0.13*    | 0.07     | 0.20***  | −0.05    | 0.84***  |          |          |          |          |          |
| **Grades**     |         |          |          |          |          |          |        |          |          |          |          |          |          |          |          |          |          |          |          |
| $Math$         | 3.1     | 1.0      | 0.15***  | 0.07     | 0.11     | <−0.001  | 0.13*  | −0.18**  | 0.30***  | 0.18**   | 0.31***  | 0.10     | 0.35***  | −0.31*** |          |          |          |          |          |
| $Phy$          | 2.8     | 0.9      | 0.13**   | 0.08     | 0.16     | 0.03    | 0.12*  | −0.12    | 0.13     | 0.10     | 0.18**   | 0.10     | 0.16     | −0.14    |          |          |          |          |          |
| $Ger$          | 3.0     | 0.7      | 0.03     | 0.02     | 0.06     | <−0.001  | 0.14**  | <−0.01   | 0.04     | 0.05     | −0.04    | 0.09     | 0.04     | −0.01    |          |          |          |          |          |

*Note. N = 319–325; internal consistencies ($\alpha$) are bold numbers in diagonal. P: Performance; M: Mastery; HS: hope for success; FF: fear of failure; Ap: Approach; Av: Avoidance; math: mathematics; Phy: physics; Ger: German. Achievement motivation items ranged from 1 to 4 and grades ranged from 1 to 6. *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$. 
Appendix 2. Adapting scales to domain-specificity in Study 1

*Adaptation of the Achievement Goal Questionnaire* (Elliot & McGregor, 2001). The four scales have three items each with a response format ranging from 1 = *strongly disagree* to 4 = *strongly agree*. We chose to use an even number of answer alternatives to avoid critical issues related to the midpoint of a rating scale (Kulas & Stachowski, 2009; Rost, Carstensen, & von Davier, 1999). To construct math-specific items, the phrase ‘in math class’ was added (e.g. ‘My goal in math class is to get a better grade than the other students’).

*Adaptation of the Revised German Achievement Motives Scale* (AMS-R; Lang & Fries, 2006; cf. Nygård & Gjesme, 1973). The revised German Achievement Motives Scale (AMS-R; Lang & Fries, 2006; cf. Nygård & Gjesme, 1973) measures the motives hope for success (HS) and fear of failure (FF) with 10 items that are answered on a 4-point scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. Math-specific items were constructed by adding ‘in math’ (e.g. ‘In math, I like it when I can find out how capable I am’).

Appendix 3. Adapting scales to domain-specificity in Study 2

*Achievement motivation: German version* (Dahme et al., 1993) of the Achievement Motives Scale (AMS) by Nygård and Gjesme (1973). The original instrument uses 14 items each to assess hope for success and fear of failure (McClelland et al., 1953). In this study a short form was applied, measuring each subscale by seven items on a 4-point Likert-type scale ranging from 1 = *strongly disagree* at all to 4 = *strongly agree*. The factor structure of the German short version has been confirmed and the questionnaire has already been applied in other empirical studies (Steinmayr & Spinath, 2008, 2009). The original items are worded in a general fashion. In order to assess domain-specific achievement motivation the original items were combined with three answer sections titled by ‘In general’, ‘mathematics’, and ‘German’. It is worth noting that this implies that the process of contextualising items was slightly different to Study 1. There, the phrase ‘in math’ was added to each of the items.

*Academic self-concept: Scales for the assessment of academic self-concept* (SESSKO; Dickhäuser, Schöne, Spinath, & Stiensmeier-Pelster, 2002; Schöne et al., 2003). The authors of the SESSKO scales defined academic self-concept as ‘all cognitive representations of one’s own abilities in academic achievement situations’, for example at school (Dickhäuser et al., 2002, p. 394). The subscale called self-concept absolute was used in Study 2. Participants rated on a five-point rating-scale how good they think they are at school in general, in mathematics and in German, respectively. General and domain-specific self-concept in math and German were assessed by four items each. The general items were taken from the described questionnaire (Dickhäuser et al., 2002; Schöne et al., 2003).

*Domain-specific self-esteem: The item ‘I have high self-esteem’ was combined with three answer sections titled by ‘In general’, ‘mathematics’, and ‘German’, to construct domain-specific self-esteem items. Each answer section had a 4-point Likert-type scale ranging from 1 = strongly disagree to 4 = strongly agree.*
### Appendix 4. Ranges, means, standard deviations, and internal consistencies for all used variables in Study 2

|                              | Range | Mean | Standard deviation | Cronbach α |
|------------------------------|-------|------|--------------------|------------|
| **Hope for success**         |       |      |                    |            |
| HS_general                   | 3.29  | 2.94 | 0.60               | 0.89       |
| HS_math                      | 3.43  | 2.62 | 0.79               | 0.92       |
| HS_German                    | 3.00  | 2.63 | 0.68               | 0.91       |
| **Fear of failure**          |       |      |                    |            |
| FF_general                   | 3.00  | 1.89 | 0.64               | 0.86       |
| FF_math                      | 3.00  | 2.01 | 0.72               | 0.90       |
| FF_German                    | 3.00  | 1.90 | 0.68               | 0.90       |
| **Scholastic performance**   |       |      |                    |            |
| math Grade                   | 5.00  | 2.94 | 1.10               | -^a        |
| German Grade                 | 50.00 | 20.87| 0.82               | -^a        |
| **Self-concept**             |       |      |                    |            |
| Sc_math                      | 4.00  | 3.22 | 1.00               | 0.95       |
| Sc_German                    | 4.00  | 3.59 | 0.82               | 0.92       |
| **Self-esteem**              |       |      |                    |            |
| Se_math                      | 3.00  | 2.57 | 0.99               | -^a        |
| Se_German                    | 3.00  | 2.72 | 0.91               | -^a        |

Note. N_{HS, FF, Sc} = 390, N_{grades} = 388, N_{Se} = 382. HS: hope for success; FF: fear of failure; Sc: Self-concept; Se: Self-esteem. The achievement motive scale and the single-item self-esteem scale range from 1 to 4, the scales for the assessment of academic self-concept and grades range from 1 to 6. ^aCronbach’s α is not reported as the variable is assessed as single-item measure.