Learning and Instruction 107 (2027) 102471

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

The underestimated role of effort in ability self-concept formation: Preliminary evidence for another major comparison process

Tobias Debatin [a,*], Heidrun Stoeger [a], Albert Ziegler [b]

[a] Department of Educational Sciences, University of Regensburg, Regensburg, Germany
[b] Department of Educational Psychology, Friedrich-Alexander University Erlangen-Nuremberg, Nuremberg, Germany

ARTICLE INFO

Keywords:
Academic self-concept
Effort
Dimensional comparisons
Social comparisons
Metacognition

ABSTRACT

Background: Early studies found that, among individuals with equal achievement, those perceived to exert less effort were rated as having greater ability (i.e. there was a negative relationship between effort and ability), leading to proposals that effort-achievement comparisons are highly influential in forming ability self-concepts. We argue that effort-achievement comparisons have been neglected in recent research due to large-scale studies failing to confirm the early findings. However, this may be because these studies have shifted their focus to a different type of effort: goal-driven effort, meaning perceived effort an individual actively chooses to invest.
Aims: This study examines whether the negative relationship between effort and ability self-concepts among individuals with equal achievement can be confirmed when effort is operationalized more similarly to the earlier studies, as so-called data-driven effort, meaning perceived effort resulting from task requirements.
Sample: Participants were 448 ninth-grade students from Germany.
Method: Data-driven effort, achievements (school grades) and ability self-concepts in the domains of STEM and languages were assessed. We used structural equation modeling to test our hypotheses.
Results: There was a strong negative relationship between data-driven effort and ability self-concepts in the two domains when controlling for achievements in the domains.
Conclusions: In line with older self-concept research, these findings suggest that there is another influential comparison process, in which data-driven effort is compared with achievements, a topic that merits further investigation in future research.

1. Introduction

Ability self-concepts are evaluative self-perceptions of abilities, which are important for many outcomes like test anxiety (Arens et al., 2017), career aspirations (Debatin et al., 2025), and school achievements (Marsh, 2023). Three major comparison processes are currently assumed to be the (proximal) main drivers of individuals' ability self-concept formation: social comparisons of achievements (i.e., external comparisons of one's achievements with the achievements of other persons; Festinger, 1954), dimensional comparisons of achievements (i.e., internal comparisons of one's achievement in one domain with achievements in other domains; Marsh, 1986; Möller & Marsh, 2013), and temporal comparisons of achievements (i.e., internal comparisons of one's current achievement with prior achievements; Albert, 1977).

However, based on attribution theories, there has been a prominent discussion about another important comparison process for ability self-concept formation, namely, a comparison of one's achievements with one's involved effort (Nicholls, 1978, 1984), which is neglected nowadays. Such a process was derived from several vignette studies, which found that when judging the abilities of two persons (comparisons of two others or oneself with another person) who showed the same achievements, but differed in the amount of effort they invested, the person who invested less effort was rated as having more ability (Muenks & Miele, 2017; Nicholls, 1978, 1984). This is in line with a compensation schema, namely that a lack of ability must be compensated by effort to reach the same performance (Kun, 1977; Muenks & Miele, 2017), which might negatively affect the self-concept. However, two more recent large-scale observational studies could not clearly confirm this negative relation between effort and ability self-concepts in students with the same achievements (Marsh et al., 2016; Pinxten et al., 2014), and, subsequently, the role of effort became somewhat neglected

* Corresponding author. Department of Educational Sciences, University of Regensburg, Universitaetsstraße 31, 93053, Regensburg, Germany.
E-mail address: tobias.debatin@ur.de (T. Debatin).

https://doi.org/10.1016/j.learninstruc.2026.102471
Received 10 November 2025; Received in revised form 20 July 2026; Accepted 3 September 2026
Available online 10 September 2026

in self-concept research.

However, based on our reading, we assume that the inconsistent findings on the relationship between effort and ability self-concepts may be due to the inconsistent way effort was conceptualized and measured. Following definitions from metacognitive research (Koriat et al., 2006), earlier studies assessed effort resulting from task requirements (i.e., data-driven effort), e.g., due to subjective difficulty, and more recent studies assessed a different kind of effort, namely, effort an individual actively chooses to invest, e.g., due to personal goals or valuing tasks (goal-driven effort). More precisely, data-driven effort is the perceived effort, which is externally triggered by tasks when working on them, while goal-driven effort is the perceived effort, which is internally triggered, usually by motivational aspects like goals (Klepsch & Seufert, 2021). For example, when working on math tasks, individuals experience, rather passively, how effortful these tasks are for them (data-driven effort), while goal-driven effort reflects individuals' decisions about how much effort they want to invest in certain tasks within a domain based on internal factors, such as their goals. The two constructs show very different correlational patterns with other constructs. For example, in two studies by Klepsch and Seufert (2021), data-driven effort showed a positive correlation of around .60 with the subjective assessment of the complexity of a task, labeled intrinsic cognitive load by the authors, while goal-driven effort only showed a correlation of around .20. Furthermore, data-driven effort showed a clear negative correlation with performance judgments in many metacognitive studies, while goal-driven effort even showed a tendency for a positive correlation (Baars et al., 2020). To differentiate these constructs from related ones, it is important to note that data-driven effort is neither equivalent to nor solely determined by perceived difficulty. Rather, it is assumed to be influenced by additional factors, such as task presentation, duration of task engagement, and worry cognitions triggered by tasks (Endres et al., 2025). Goal-driven effort shares similarities with effort cost as conceptualized in situated expectancy-value theory (Eccles & Wigfield, 2024). However, unlike effort cost (Flake et al., 2015), it does not refer to whether a task is worth the effort; instead, it reflects the more quantitative and neutral perceived amount of effort an individual chooses to invest.

The primary goal of our study was to examine whether the negative relationship between effort and ability self-concepts can be confirmed when effort is operationalized more similarly to the earlier studies as data-driven effort. First, we will provide a brief overview of the major comparison processes involved in ability self-concept formation and the role of effort in shaping ability self-concept. We will argue that recent self-concept research is not paying enough attention to effort in comparison processes. Second, we will review the somewhat unclear findings regarding the role of effort in the formation of ability self-concept, to clarify why its role seems currently to be underestimated. Finally, we will outline why data-driven effort seems crucial in self-concept formation and present empirical evidence that it indeed plays a vital role.

1.1. *The major comparison processes of ability self-concept formation and the role of effort*

Marsh (1986) proposed the prominent internal/external frame of reference model (I/E Model), which assumes that ability self-concepts are formed via specific external and internal frames of reference for achievement comparisons. First, social (external) comparisons of one's achievements in a subject with those of others in the same subject are important for the formation of self-concepts; that is, the better a person's achievements in a subject are compared to those of others, the higher their self-concepts. Second, intra-individual (internal) comparisons between different subjects play a role; meaning that the better a person's achievements in, e.g., mathematics, are relative to another subject, e.g., languages, the lower their self-concept in languages. Initially, research focused on mathematics and verbal achievements for internal comparisons, but over time, it was extended to other achievement domains, culminating in dimensional comparison theory (Möller & Marsh, 2013). A different type of internal achievement comparison is a temporal comparison process (Albert, 1977), in which one's current achievements are compared with prior achievements. This process was accepted as a third major comparison process in ability self-concept development (Skaalvik & Skaalvik, 2002; Wilson & Ross, 2000) and was integrated in the later 2/IE model (Wolff, Nagy, et al., 2019) and its extensions (Wolff et al., 2021; Wolff, Helm, & Möller, 2019).

In the 1980s and 1990s, there was also discussion about a fourth major comparison process for ability self-concept formation, which involves effort and achievement (Nicholls, 1984; Skaalvik & Skaalvik, 2002). Nicholls (1984) suggested that "to judge our capacity we must compare the effort and attainment of self and others" (p. 329) and Skaalvik and Skaalvik (2002) also emphasized the importance of comparisons that involve both effort and achievement. Nicholls (1984) quite clearly refers to effort resulting from task requirements, very similar to the mentioned descriptions of data-driven effort: "effort or time one needs to learn something (compared to the effort or time it takes others)" (p. 329), while the wordings of Skaalvik and Skaalvik (2002) are less clear regarding what kind of effort they describe. However, both articles suggest a central internal comparison process between effort and achievement in the formation of domain-specific ability self-concepts: the higher a person's perceived effort is in a domain (e.g. judged to be high relative to classmates due to observations and communication) compared to their achievements in a domain (e.g. judged to be average relative to classmates), the lower their self-concept is. As shown in our examples and according to Skaalvik and Skaalvik (2002), such internal comparison processes of students can be based on external social comparisons. For example, when the perceived relative ranks of achievements within a class in two subjects are used as a basis for dimensional comparisons. We believe that this may also be the case when comparing achievement with effort in the same subject, meaning that effort is also perceived as low, average or high relative to the observed effort of classmates. As another example, students could have average achievement in math relative to their classmates, but low perceived effort relative to them, resulting in a self-concept that is higher than would be expected based on achievement alone. Additionally, Skaalvik and Skaalvik (2002) suggested another comparison process relating to the effort required to achieve in different subjects, i.e., a dimensional comparison of effort. They assumed that individuals have a higher self-concept in subjects in which they invest less effort to reach the same level of achievement.

In summary, it appears that comparisons between effort and achievement are central to the formation of ability self-concept but have not received the necessary attention they deserve in current research. This may be due in part to somewhat unclear empirical findings, which we will summarize below.

Several early vignette studies found that when judging the abilities of two persons (comparisons of two others or oneself with another person; the latter being an ability self-concept assessment), who showed the same performance (number of correct answers), but differed in the amount of effort (time actively working on tasks), the person who invested less effort was rated as having more ability (Muenks & Miele, 2017; Nicholls, 1978; Nicholls & Miller, 1984). There is consensus that in such scenarios and from a certain age on, almost all individuals perceive ability and effort to be negatively related (Muenks & Miele, 2017), probably due to the assumption that a lack of ability has to be compensated by effort to reach the same performance (Kun, 1977). These early findings from vignette studies indicate that people compare effort with achievement to judge how good or talented someone is. Based on these findings, it is reasonable to infer that effort assessments affect self-concepts.

However, in more recent large-scale observational studies, researchers have only found very small, barely significant negative effects (Marsh et al., 2016), or no relations at all (Pinxten et al., 2014) between effort and ability self-concepts in students with the same level of achievement. We argue that the lack of significant effects in very large samples, together with the much lower effect sizes than the effects of achievements in the same studies (which were about four to 10 times the size), led to the assumption that effort plays a less prominent role in ability self-concept formation than earlier studies suggested. Accordingly, the topic was subsequently neglected in research on ability self-concepts. In the next section, we will argue that the type of effort determines its importance for self-concept formation and that the low or absent effects of effort in the two large-scale studies are due to the fact that a different type of effort was investigated than in the earlier vignette studies.

1.2. Two different types of effort and their role in ability self-concept formation

Research in the area of metacognition may help better understand the differences in the types of effort examined in earlier and recent self-concept research, thereby aiding understanding of inconsistent findings. There is an important distinction between at least two types of effort based on its origin (Baars et al., 2020; Koriat, 2018; Koriat et al., 2006): perceived effort an individual actively chooses to invest in reaching goals or due to valuing tasks, so-called goal-driven effort (involving self-agency), and perceived effort resulting from task requirements, like the subjective difficulty of a task, so-called data-driven effort. Research has shown that these two types of effort are associated to varying degrees with performance judgments (Koriat, 2018). It appears that effort, driven by task requirements (i.e., data-driven effort), is more likely to lead to negative performance judgments (Koriat et al., 2006). The more effort people require to solve tasks in a domain, the worse they judge their performance on those tasks and in the domain. In contrast, effort that an individual actively chooses to invest because of the value of a task (i.e., goal-driven effort), instead of being mainly the result of task requirements and the subjective difficulty of the task (i.e., data-driven effort), seems to be associated with effective learning behavior and skill development (Grund et al., 2024). In turn, goal-driven effort can lead to more positive performance judgments.

It is important to note that the performance judgments in these metacognitive studies are not equivalent to ability self-concepts. Usually, so-called judgments of learning, confidence ratings, or ease-of-learning judgments are used (Baars et al., 2020). They can be prospective, concurrent (immediately after answering), or more retrospective judgments, and can be more item-specific or more global (Baars et al., 2020). However, many of these judgments are very similar to operationalizations of self-efficacy and ability self-concept (which are highly correlated, though distinguishable (Marsh et al., 2019)). Research on performance judgments might, thus, provide important insights into the role of different types of effort and their relationships to ability self-concepts. In particular, studies have found that the amount of data-driven effort is negatively related to performance judgments, similar to early studies on self-concept formation, whereas goal-driven effort is positively related to or unrelated to performance judgments, similar to recent self-concept studies (for a meta-analysis, see Baars et al., 2020).

Applying this distinction between two types of effort, we argue that, in contrast to the vignette studies mentioned earlier, the two more recent large-scale studies on self-concept used operationalizations of goal-driven effort (i.e., actively chosen and likely goal-directed investment of effort), rather than data-driven effort (i.e., effort due to task or domain requirements). For example, Marsh et al. (2016) used items such as "In math, I invest much effort to understand everything," and "In math, I try my best to do everything as well as possible," which refer to actively chosen and goal-directed effort, similar to the operationalization used by Koriat et al. (2018). The active effort asked for in these items is likely a consequence of value beliefs like the importance someone places on the task (in this case, a domain), as described in Grund et al. (2024) and Koriat et al. (2006) for goal-driven effort. Pinxten et al. (2014) used the two items: "I put a lot of effort into mathematics," and "I work hard for mathematics." This is also consistent with descriptions of goal-driven effort as "effort that is due to their own self-agency" (Koriat, 2018, p. 371), and these wordings are basically the same as in metacognitive studies (for an overview, see Endres et al., 2025), except that they are domain-specific instead of task-specific. Such active goal-driven effort, which seems to result from motivational sources, should be very different from data-driven effort that results mainly from the task or domain requirements. The effort required to solve tasks in a domain (data-driven effort) should be more important in shaping ability self-concepts than the effort individuals choose to invest, for example, simply because they like the tasks in that domain. Such goal-driven effort does not necessarily provide information about individuals' abilities but rather about their motivation, which could stem from many sources and thus should not influence ability self-concepts in the same way.

Overall, given the differences between data-driven and goal-driven effort, and the way large-scale studies operationalized and measured effort, or in other words, the fact that these studies investigated goal-driven instead of data-driven effort, it seems plausible that these studies did not find the negative associations between effort and self-concept expected from the vignette studies reported above. Accordingly, studies evaluating the importance of data-driven effort are clearly needed. Unfortunately, in metacognitive research, single items were usually used to assess data-driven effort, and they were typically task-specific.[1] We think one possibility for assessing data-driven effort in a domain with a scale, despite being developed in the context of situated expectancy-value theory (Eccles & Wigfield, 2024), is the scale "effort required" from Gaspard et al. (2015), one item of which reads "Engaging with math is effortful."[2] In this scale, there is no active, self-agentic aspect; instead, effort is depicted as a consequence of engaging with tasks in the domain. This is consistent with the description of data-driven effort as rather passive, resulting from task requirements rather than active choice, and with the wording of previous single items in metacognitive research (Endres et al., 2025; Koriat, 2018).

We also took a close look at other standard scales from situated expectancy-value theory that assess aspects of "effort cost" (for an overview, see Flake et al., 2015). However, all of these scales are similar to goal-driven effort but not to data-driven effort, with the "effort required" scale from Gaspard et al. (2015) being the exception. This is because effort cost, as defined and typically operationalized in situated expectancy-value theory, involves a motivational and self-agency aspect. This can be seen in the second part of the recent definition as "individuals' perceptions of how much effort they would need to exert to complete a task and whether it is worth doing so" (Eccles & Wigfield, 2024, p. 5). Accordingly, none of the scales were suitable as they typically also involve a motivational, goal-driven component by definition. Even the newer and more similar scale, "task effort cost," aims to "capture a cost and barrier to motivation" (Flake et al., 2015, p. 237) by using the phrase "too much effort" rather than simply assessing the required amount of effort.

1.3. Current study

While several older vignette studies suggest that effort plays an important role in forming ability self-concepts and indicate a negative

[1] Please note that recently and after conducting our study a new scale assessing data-driven and goal-driven effort was published (Endres et al., 2025). However, the authors opted for the terms "mental load" and "mental effort" as these were more idiomatic to the participants in their qualitative interviews than the also discussed and more original terms "data-driven effort" and "goal-driven effort", which we clearly prefer.

[2] We think this is the most literal translation of the original German item "Mich mit Mathe zu beschäftigen ist anstrengend".

relation between effort and ability self-concepts among students with the same achievements, more recent large-scale observational studies have not confirmed these earlier findings. We suggest that the discrepancies can be attributed to the fact that more recent studies assessed a different kind of effort, namely, goal-driven effort (instead of data-driven effort as used in the older vignette studies). The main aim of our study was, therefore, to test whether we could confirm the negative relation between effort and ability self-concepts in students with the same achievements in an observational study, when effort was operationalized more similarly to earlier studies as data-driven effort (perceived effort due to task or domain requirements). In line with Nicholls (1984) we assume that "effort or time one needs to learn something" (p. 329), in other words, data-driven effort, is the basis for an underlying internal comparison process between perceived effort and achievement in this domain: the higher a person's perceived effort is in a domain (e.g., judged to be high relative to classmates due to observations and communication) compared to their achievements in a domain (e.g. judged to be average relative to classmates), the lower their self-concept. In this context, we tested a primary and a secondary hypothesis.

Our *primary hypothesis* was that data-driven effort in STEM/languages has an incremental negative effect on the corresponding STEM/languages self-concepts, after controlling for STEM achievement and language achievement. We assume that the effects of effort reflect a comparison between effort and achievement, while the effects of the achievement variables represent the classic social and dimensional comparison effects of achievements as interpreted in previous studies. In contrast to previous studies, we expect an effort effect size similar to that of the achievement variables (rather than much smaller) due to our different operationalization as data-driven effort.

Additionally, based on the suggestion of Skaalvik and Skaalvik (2002) that persons have a higher self-concept in subjects in which they invest less effort to reach the same level of achievement, i.e., a dimensional comparison effect of effort, we also tested a *secondary hypothesis*: There are positive effects of data-driven effort in STEM/languages on the self-concepts in the non-corresponding domain (languages/STEM), when controlling for achievements. These cross-domain effects of effort are assumed to represent dimensional comparison effects of the effort variables.

2. Method

2.1. Sample and procedure

The sample consists of 448 students from Germany (55.1% boys, 43.5% girls, 1.3% diverse; mean age: $M = 15.32$ years, $SD = .45$; 30% with a migration background) from 28 classes of 12 "Gymnasien" (the highest level secondary school in the German tracking system) in four German federal states. Students were in grade 9, the study was conducted in May and June 2023, six to eight weeks before the end of the school year. Regarding sample size, we collected as many observations as we could in the participating classes. Overall, there were 4.8% missing values. We have complied with the APA ethical standards in the treatment of our sample and in describing the details of treatment; students' and parents' informed consent was obtained for each participant. The responsible ministries of education had approved the conduct of the study in advance. The same sample was part of a previous study with completely different research questions in the context of a larger research project (Pfeuffer et al., 2025). The study and the analysis plan were not preregistered. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

2.2. Measures

2.2.1. STEM and languages self-concepts

We measured students' STEM and language self-concepts by adapting Dweck and Henderson's (1988) scale of self-confidence in intelligence to these domains. Item endpoints were changed as follows for the items: Item 1: "I wonder if I am talented for the STEM/language subjects." vs. "On the whole, I think I am talented for the STEM/language subjects.". Item 2: "When I get new learning material in the STEM/language subjects, I often think I may not be able to understand it" vs. "When I get new learning material in the STEM/language subjects, I am usually able to understand it.". Item 3: "I am not very confident about my abilities for the STEM/language subjects." vs. "I feel fully confident about my abilities for the STEM/language subjects.". A 6-point Likert-type scale was used with these contrasting statements as the endpoints. Additionally, we added a fourth item: "I am not sure if I am good enough to succeed in the STEM/language subjects." vs. "I am sure that I am good enough to succeed in the STEM/language subjects.".

While statements about being talented and understanding are typical for self-concept items (e.g., Marsh et al., 2016, 2019), other wordings of the scale might also be classified as measuring generalized self-efficacy or success expectancies. However, it was found that measures of generalized self-efficacy as well as success expectancies are empirically close to indistinguishable from (academic) self-concept (Marsh et al., 2019). In line with these findings, the measurement models of our study confirmed unidimensionality of the self-concept scales with high standardized loadings (ranging from .74 to .95). Additionally, Schroeders and Jansen (2020) showed that the global science self-concept scale used in PISA is close to equivalent to an aggregation of subject-specific self-concepts in biology, chemistry and physics, validating the use of global scales as in our study. Cronbach's alphas for the scales were .91 (STEM) and .95 (languages).

2.2.2. STEM and language effort (data-driven)

We measured data-driven effort with two items of the scale "effort required" from Gaspard et al. (2015), adapted to the domains of STEM and languages ("Engaging with the STEM/language subjects is effortful, " and "The engagement with the STEM/language subjects costs me a lot of energy."). We decided against using the two other items of the scale for our questionnaire as both seemed too emotionally negative for our purpose (especially one of them, which uses "feel completely drained after", but also the other, which uses "exhausts me"). A 6-point Likert-type format was used with endpoints ranging from "not true at all" to "completely true". Cronbach's alphas for the scales were .87 (STEM) and .91 (languages).

2.2.3. STEM and language achievement

We measured STEM and language achievement with self-reported grades from the most recent report card in mathematics, physics, chemistry, biology, computer science, German, and the first foreign language (L2). For self-reported grades in this age group, high accuracy and reliability between .80 and .90 can be assumed (Dickhäuser & Plenter, 2005; Möller et al., 2006) and they may even be more suitable for self-concept studies as students should use the grades they remember for comparisons (Wolff et al., 2021). Grades in Germany range from 1 = excellent to 6 = insufficient. We reverse-coded students' grades, with higher scores indicating higher achievements. For our final models, we used the mean of the five STEM subjects as our composite STEM achievement variable and the mean of the two language grades as our composite language achievement variable.

2.3. Data analysis

We used structural equation modeling to test our hypotheses following the refined procedure suggested by Marsh et al. (2015) for I/E models: The STEM and languages self-concepts were modelled as two separate latent factors. As the items in the two domains had the same wording, we specified correlated residuals between the corresponding items. In Model 1, both self-concept factors were regressed only on the achievement mean scores in the two domains for calculating the typical social and dimensional comparison effects. In Model 2, we added the effort mean scores in the two domains as additional predictors. We included the effort mean scores also in Model 1, but only as correlated variables, not affecting the regression estimates. Using this approach, the two models have the same degrees of freedom and model fit, which allows a direct comparison regarding different sets of predictors (Marsh et al., 2015). The two models also included the covariates gender, home language (German vs. other), SES, and age.

In the following, we want to explain why we decided to use mean scores instead of latent variables for the predictors in our final models. First, while a reflective measurement model would make sense for the effort variables, it does not for the grades in the school subjects, as we do not assume latent variables that are the (single) causes of the grades in all STEM subjects or language subjects. Therefore, mean scores should be used as otherwise distortions of the meaning and structural parameters would arise (Rhemtulla et al., 2019). As the (composite) mean scores of the grades contain measurement error, another problem could result, in case we would model the effort variables as latent variables: We would use error-free variables as well as variables with measurement error as predictors, which would probably lead to positively biased estimates of the paths of the effort variables and negatively biased paths of the achievement variables (Cole & Preacher, 2014). Considering that all mean score variables would likely contain very little measurement error (the effort variables showed very high reliabilities (.87 and .91) and for the self-reported grades high reliability between .80 and .90 can also be assumed (Dickhäuser & Plenter, 2005; Möller et al., 2006), we decided it would be better to accept that all paths are probably slightly negatively biased when using mean scores (Cole & Preacher, 2014).

Following the recommendations in a recent editorial of the executive director of the American Statistical Association (ASA) and colleagues (Wasserstein et al., 2019), we will not use the term "statistically significant" or a particular threshold for p values like $<.05$ in this manuscript. Instead, we will treat p values "as a continuous measure of the compatibility between the data and the entire model used to compute it", as supported by the ASA (Greenland et al., 2016, p. 339). The term "entire model" is used to highlight the fact that p values do not solely depend on how compatible the data are with a test hypothesis (often a null hypothesis). Instead, violations of a host of other assumptions about data generation, such as a properly conducted analysis, proper measurement of constructs, and adherence to the study protocols, may impact p values (Amrhein et al., 2019; Greenland et al., 2016). In other words, any violation of the aforementioned assumptions could lead to low or high p values. Accordingly, the point estimates of the effects are the effect sizes most compatible with the data (they have $p = 1$) if all other assumptions are correct. Evidence for or against hypotheses will be judged by considering a combination of precise p values (except when $p < .001$), effect sizes, standard errors, consistency with theory and prior results, and factors like measurement quality. Additionally, as supported by the American Statistical Association, we used one-sided p values for our one-sided hypotheses (point 14 in Greenland et al., 2016).

2.4. Estimation of the models

We used R 4.3.1 (R Core Team, 2023) with the lavaan package (v0.6-16; Rosseel, 2012) for the main analyses. We also used the DataExplorer package (v0.8.2; Cui, 2020) for getting our missing data profile and the distributions of the variables. The robust maximum likelihood estimator (MLR) with the cluster command of lavaan was used for all analyses to appropriately adjust the standard errors as students were clustered within classes. Model fit was assessed following the criteria of Hu and Bentler (1999). Therefore, a value close to .95 for the Comparative Fit Index (CFI), a value close to .06 for the root mean squared error of approximation (RMSEA), and a value close to .08 for the standardized root mean squared residual (SRMR) were the cutoff criteria for assuming good model fit. Missing values were handled using the full information maximum likelihood method (FIML).

The analysis code for the main analyses of this study can be found in the online supplemental material. Research data are openly available via the link at the end of the manuscript.

3. Results

As a reminder, we will not use the term "statistically significant" or a particular threshold for p values like $<.05$ in this manuscript. Instead, we will treat p values "as a continuous measure of the compatibility between the data and the entire model used to compute it", as supported by the ASA (Greenland et al., 2016, p. 339) and described in the data analysis section.

Table 1 shows the descriptive statistics and correlations for all variables. Most importantly, data-driven effort showed rather high negative correlations with self-concept in both domains (−.52 (languages) and −.54 (STEM)).

3.1. Model fit

As discussed before, the model fit of the two main models is identical, and the indices indicated very good model fit (χ^2scaled(69) = 115.31, $p < .001$; CFI(robust) = .99, RMSEA(robust) = .04, SRMR = .02).

3.2. Primary hypothesis: incremental negative effects of data-driven effort in the same domain

Note: As supported by the American Statistical Association, we used one-sided p values for our one-sided hypotheses (point 14 in Greenland et al., 2016). As reported in Table 2 and depicted in Fig. 1, we found evidence in Model 1 (I/E-Model) for the classic social and dimensional comparison effects, namely positive effects of the achievement variables on the self-concepts in the same domain and negative effects of the achievement variables on the self-concepts in the other domain. Most importantly, in Model 2 we also found evidence for the primary hypothesis of substantial incremental negative effects of data-driven effort on the self-concepts in the same domain (STEM: −.36, $p < .001$, one-sided, 95% CI [−.44, −.27]; Languages: −.38, $p < .001$, one-sided, 95% CI [−.46, −.30]). Complete results including the effects of the covariates can be found in the supplementary materials.

3.3. Secondary hypothesis: dimensional comparison effects of data-driven effort

We did not find clear evidence in Model 2 for the expected

Table 1
Descriptive statistics and correlations.

	1	2	3	4	5	6
1. STEM self-concepts	—					
2. STEM achievement	.54*	—				
3. STEM effort	−.54*	−.36*	—			
4. Languages self-concepts	.04	.14*	.03	—		
5. Languages achievement	.15*	.57*	−.11*	.50*	—	
6. Languages effort	.06	.00	.06	−.52*	−.31*	—
M	3.66	4.56	3.59	4.01	4.67	3.37
SD	1.33	.80	1.26	1.37	.79	1.26

Note: $N = 448$. The values are extracted from the final model. The descriptive statistics of the self-concepts in STEM and languages are based on their observed means as latent means are only relative. All variables ranged from 1 to 6, with 6 indicating the highest. * = $p < .05$.

Table 2
Standardized path coefficients of the two models.

Predictors	STEM self-concepts					
	Model 1 (I/E)			Model 2 (I/E + effort)		
	Est.	*SE*	*p*	Est.	*SE*	*p*
STEM achievement	.59	.05	<.001	.44	.06	<.001
Languages achievement	−.14	.06	.007	−.09	.06	.073
STEM effort	-	-	-	−.36	.04	<.001
Languages effort	-	-	-	.03	.04	.246
Explained variance	37.5%			47.9%		
Predictors	**Languages self-concepts**					
	Model 1 (I/E)			Model 2 (I/E + effort)		
	Est.	*SE*	*p*	Est.	*SE*	*p*
Languages achievement	.63	.06	<.001	.45	.06	<.001
STEM achievement	−.22	.06	<.001	−.10	.05	.024
Languages effort	-	-	-	−.38	.04	<.001
STEM effort	-	-	-	.08	.05	.074
Explained variance	29.4%			41.5%		

Note: $N = 448$. *P* values are one-sided due to our one-sided hypotheses. All models include the covariates gender, home language (German vs. other), SES, and age.

dimensional comparison effects of the effort variables on self-concepts in the other domain, meaning positive effects of data-driven effort in one domain on the self-concepts in the other domain (STEM: .03, $p = .246$, one-sided, 95% CI [−.05, .12]; Languages: .08, $p = .07$, one-sided, 95% CI [−.03, .18]). However, both effects were in the expected direction, and at least for languages, the *p* value was rather low.

4. Discussion

Based on attribution theories, several older vignette studies found that effort plays an important role in forming ability self-concepts (for a review see Muenks & Miele, 2017) and there was also prominent discussion about an important comparison process of one's achievements with one's involved effort (Nicholls, 1984; Skaalvik & Skaalvik, 2002). However, more recent observational large-scale studies did not confirm these findings (Marsh et al., 2016; Pinxten et al., 2014) and the role of effort in self-concept formation was subsequently rarely discussed. We outlined that the discrepancies in findings might be because more recent studies assess a different kind of effort, namely goal-driven effort, effort an individual actively chooses to invest due to goals or valuing the task (Koriat et al., 2006), while the older studies assessed rather data-driven effort, meaning effort due to task or domain requirements (Koriat et al., 2006). Next to introducing this important differentiation to self-concept research, the main aim of our study was to test if we can find evidence for the importance of perceived effort in self-concept formation for the first time in an observational study when effort is operationalized more similarly to the earlier studies as data-driven effort. Overall, our results provide evidence for the importance of data-driven effort in the formation of ability self-concepts. However, note that, due to the cross-sectional nature of our data, we are essentially unable to find evidence for causal hypotheses. Accordingly, the word "effect" in describing our results is only "technical" because it is common to describe relations in multiple regression and structural equation models with this wording.

4.1. Correlational results

First, it is noteworthy that data-driven effort showed rather high negative correlations with self-concept in both domains (−.52 (languages) and −.54 (STEM)), while in previous studies the cross-sectional correlations of self-concepts with effort, which we classified as goal-driven effort, ranged from −.04 to .38, with an average of .23 (Marsh et al., 2016; Pinxten et al., 2014). We consider this to be evidence that the operationalization of effort certainly matters. However, we did not assess goal-driven effort in our study, so a direct comparison of the two kinds of effort in a single study was not possible with our data.

4.2. Primary hypothesis: incremental negative effects of data-driven effort in the same domain

We found evidence for our primary hypothesis that data-driven effort in STEM/languages has a large incremental negative effect on STEM/languages self-concepts, when controlling for STEM achievement and languages achievement. This means that when people show the same level of achievement in a domain, the higher they perceive their required effort in this domain, the lower their corresponding ability self-concepts are, replicating findings from earlier vignette studies (Muenks & Miele, 2017; Nicholls, 1978; Nicholls & Miller, 1984). Thus, findings

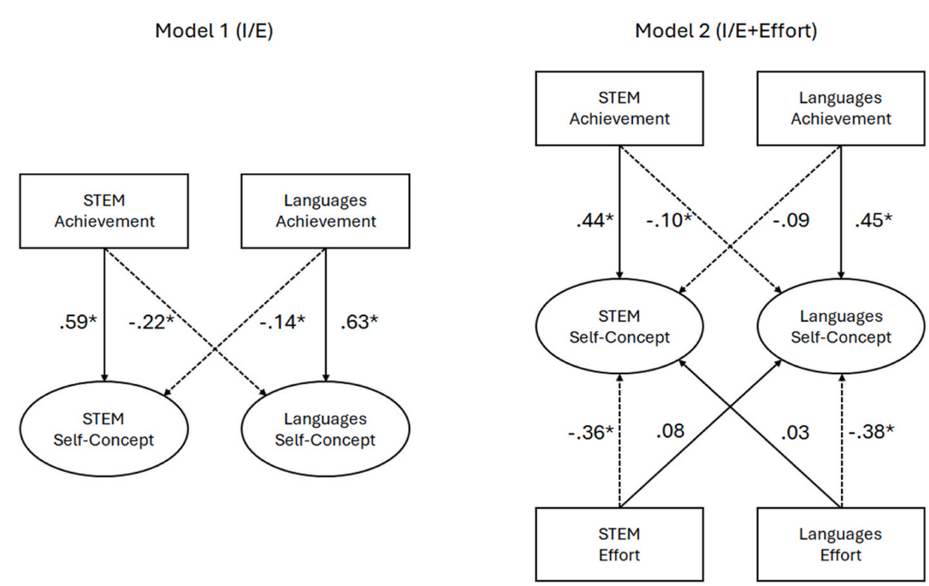

Fig. 1. Simplified Depiction of the two Models.
Note. Hypothesized positive effects are indicated by continuous lines and hypothesized negative effects by dashed lines. Path coefficients are standardized. * indicates $p < .05$.

from metacognitive research, in which data-driven effort is typically negatively related to performance judgements (Baars et al., 2020), seem to generalize to relations with ability self-concepts, also and importantly when controlling for achievement and when using a domain-instead of task-specific operationalization of data-driven effort. Noteworthy, adding data-driven effort variables as predictors increased the explained variance from 37.5% in a model including only the achievement variables to 47.9% for STEM self-concepts, and from 29.4% to 41.5% for languages self-concepts. These findings indicate that these variables certainly deserve more attention.

Based on prior research and as described in the introduction, we think the most plausible reason for this finding is that another important comparison process is involved in forming ability self-concepts, namely a comparison with an internal frame of reference between required effort and achievement in the same domain, which leads to high self-concepts when there is the perception of succeeding without much effort (Skaalvik & Skaalvik, 2002). This likely also involves external comparisons with the effort and achievement of other people as a source of information in such internal comparisons, illustrating the complexities of such processes (Skaalvik & Skaalvik, 2002). Accordingly, it would also be interesting to use scales that ask explicitly for such external social comparisons of effort, like Lee et al. (2025) did, but focusing specifically on data-driven effort.

We want to note that we disagree with Grund et al. (2024) that data-driven effort (respectively, effort-by-complexity as they call it) is closely related to effort cost from situated expectancy-value theory (Eccles & Wigfield, 2020). Effort cost, as defined and typically operationalized in situated expectancy-value theory, involves a motivational, self-agency aspect as can be seen in the second part of the recent definition as “individuals' perceptions of how much effort they would need to exert to complete a task and whether it is worth doing so” (Eccles & Wigfield, 2024, p. 5). Accordingly, for example, Battle and Wigfield (2003) used items with a motivational, goal-driven component, such as “Getting a graduate degree sounds like it really requires more effort than I'm willing to put into it”. As mentioned in the introduction, even the newer and arguably more similar scale “task effort cost” aims to “capture a cost and barrier to motivation” (Flake et al., 2015, p. 237) by intentionally using the phrase “too much effort” instead of assessing more simply the amount of required effort. These operationalizations are very different from the items we used from the scale “effort required” (Gaspard et al., 2015) to measure data-driven effort, despite also being developed in the context of situated expectancy-value theory.

4.3. Secondary hypothesis: dimensional comparison effects of data-driven effort

Regarding our secondary hypothesis on the dimensional comparison effects of effort, we found no evidence of positive effects of data-driven effort in STEM/language on self-concepts in the non-corresponding domain (languages/STEM) in our models, consistent with the null hypothesis. However, there was a tendency for such effects, at least for language self-concept. This cautiously indicates that when individuals exert the same level of effort in the domain of language, the higher they perceive their required effort in the domain of STEM, the higher their ability self-concepts are in the domain of languages. One reason for the weak effects could be that students focus on different domains when asked about STEM (Stoeger et al., 2025). Based on previous research, we know that dimensional comparison effects are most pronounced when the subjects involved are dissimilar on the mathematics-verbal continuum, meaning that, for example, comparisons between biology and languages show smaller negative effects than comparisons between mathematics and languages (Möller et al., 2020). If the students in our sample associated STEM most strongly with biology, the minor effects could result from these associations. However, as we found evidence for relatively strong traditional dimensional comparison effects in our study, and these traditional achievement-based effects have been found within the domains of STEM and languages before (Debatin et al., 2023), we assume that dimensional comparisons of data-driven effort are not a major factor influencing ability self-concepts.

4.4. Implications

Our results suggest several theoretical and practical implications. First, the large effects in the structural equation model, beyond those of the achievement variables, indicate that data-driven effort is a significant factor in explaining the formation of ability self-concepts. Accordingly, students’ perceptions of a subject as "effortful" (i.e., requiring substantial mental energy) may represent an underappreciated pathway through which ability self-concept is shaped, even among students performing equally well. The effects in the model are even larger than the established dimensional comparison effects of achievements from different domains. Together with earlier findings from vignette studies, our findings suggest that there may be another major comparison process between data-driven effort and achievement that should not be neglected. Should such an influential comparison process be confirmed in future studies, it would also be a promising new target for interventions aimed at improving ability self-concepts. For example, explicitly addressing the important role of effort in skill development and that everyone has to invest effort to improve might lead to a less negative association between hard work and abilities.

4.5. Limitations and future research

While the study provides relevant evidence for the arguably large role of data-driven effort in ability self-concept formation, there are also limitations to be kept in mind. First, our approach to operationalizing data-driven effort is only one of many possibilities. The name of the scale we used to measure data-driven effort is “effort required” and asks, for example, how effortful it is to engage in specific school subjects (Gaspard et al., 2015). However, items that directly refer to the required effort a student must invest in a domain or school subject might be an even better way to measure data-driven effort. In future studies, different operationalizations of data-driven effort could be compared for consistency of findings. Furthermore, although self-report measures are necessary for measuring perceived effort, incorporating other, more objective data sources, such as log data or physiological measures, would provide a more comprehensive understanding of the relationship between effort and self-concept of ability.

Second, data-driven and goal-driven effort were not both assessed and directly compared in our study. While this was not necessary to test the hypothesized importance of data-driven effort, it would be an important objective for future studies, among others, to find out if both kinds of effort uniquely contribute to academic self-concept formation. Comparisons with previous studies are also limited to some extent as these studies did not explicitly measure goal-driven effort but used measures which we classified as (domain-specific) goal-driven effort based on very similar wordings used in metacognitive research and definitions as outlined in the introduction.

Further, while Schroeders and Jansen (2020) showed that the global science self-concept scale used in PISA is close to equivalent to an aggregation of subject-specific self-concepts in biology, chemistry and physics, validating the use of global scales as in our study to a certain degree, they also gave an overview of research showing differentiation within the domain of science to some extent, despite a hierarchical structure. Such differentiation could affect correlational patterns, usually reducing correlations.

Another interesting consideration is how well both constructs are measurable at the domain-specific level. As goal-driven effort should reflect more stable motivational tendencies to invest effort, no matter the specific tasks, a domain specific level seems appropriate. Data-driven effort might be more task-specific and situational, but regarding effects on broader ability self-concepts, a broader aggregate of many task experiences seems most influential and, accordingly, also very suitable. However, it would be interesting to measure data-driven effort across several different tasks within a domain. Then, the average of these effort measures could be compared with domain-specific measures. Additionally, one could investigate whether the variance from task to task within a domain is also related to ability self-concepts.

Finally, similar to the other major comparison processes, in future research, the effect of data-driven effort in self-concept formation should be investigated with different study designs. Experimental, diary, and longitudinal studies could get evidence about causality, potential reciprocal effects, and the involved processes to some extent, which was basically not possible with our cross-sectional, correlational design. Additionally, the findings should be replicated in other school tracks, age groups, and educational systems.

5. Conclusion

Overall, the large incremental effects of data-driven effort on ability self-concepts in the domains of STEM and languages in our model, over and above the well-known social and dimensional comparison effects of achievements, indicate an important role of data-driven effort in the development of ability self-concepts. In line with older self-concept research, we think there is a fourth major comparison process, in which data-driven effort is compared with achievements, and which should be more closely investigated in future research.

CRediT authorship contribution statement

Tobias Debatin: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Heidrun Stoeger:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Albert Ziegler:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Data availability

The data underlying this article are openly available at https://doi.org/10.23668/psycharchives.22458.

Funding

The project was funded by the Federal Ministry of Education and Research (BMBF) under the project grant numbers 16MF1091A and 16MF1091B.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2026.102471.

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review, 84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician, 73*(sup1), 262–270. https://doi.org/10.1080/00031305.2018.1543137

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology, 109*(5), 621–634. https://doi.org/10.1037/edu0000163

Baars, M., Wijnia, L., de Bruin, A., & Paas, F. (2020). The relation between students' effort and monitoring judgments during learning: A meta-analysis. *Educational Psychology Review, 32*(4), 979–1002. https://doi.org/10.1007/s10648-020-09569-3

Battle, A., & Wigfield, A. (2003). College women's value orientations toward family, career, and graduate school. *Journal of Vocational Behavior, 62*(1), 56–75. https://doi.org/10.1016/S0001-8791(02)00037-4

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*(2), 300–315. https://doi.org/10.1037/a0033805

Cui, B. (2020). Data explorer: Automate data exploration and treatment. *R package version 0.8.2*. https://CRAN.R-project.org/package=DataExplorer.

Debatin, T., Stoeger, H., & Ziegler, A. (2023). Modeling social, dimensional, and temporal comparisons in self-concept development with the random intercept cross-lagged panel model: A methodological-substantive integration. *Developmental Psychology, 59*(9), 1595–1607. https://doi.org/10.1037/dev0001568

Debatin, T., Stoeger, H., & Ziegler, A. (2025). Are competence beliefs or value beliefs more important for STEM career aspirations?—longitudinal mediation analyses based on recent modeling approaches show different results. *Journal of Youth and Adolescence*. https://doi.org/10.1007/s10964-025-02162-3

Dickhäuser, O., & Plenter, I. (2005). Letztes Halbjahr stand ich zwei. *Zeitschrift für Padagogische Psychologie, 19*(4), 219–224. https://doi.org/10.1024/1010-0652.19.4.219

Dweck, C. S., & Henderson, V. L. (1988). *Theories of intelligence: Background and measures* (Unpublished manuscript).

Eccles, J. S., & Wigfield, A. (2024). The development, testing, and refinement of eccles, wigfield, and colleagues' situated expectancy-value model of achievement performance and choice. *Educational Psychology Review, 36*(2), 51. https://doi.org/10.1007/s10648-024-09888-9

Endres, T., Bender, L., Sepp, S., Zhang, S., David, L., Trypke, M., … Paas, F. (2025). Developing the mental effort and load–translingual scale (MEL-TS) as a foundation for translingual research in self-regulated learning. *Educational Psychology Review, 37*(1), 5. https://doi.org/10.1007/s10648-024-09978-8

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology, 107*(3), 663–677. https://doi.org/10.1037/edu0000003

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Grund, A., Fries, S., Nückles, M., Renkl, A., & Roelle, J. (2024). When is learning "Effortful"? Scrutinizing the concept of mental effort in cognitively oriented research from a motivational perspective. *Educational Psychology Review, 36*(1), 11. https://doi.org/10.1007/s10648-024-09852-7

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. https://doi.org/10.1080/10705519909540118

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education, 6*, Article 645284. https://doi.org/10.3389/feduc.2021.645284

Koriat, A. (2018). Agency attributions of mental effort during self-regulated learning. *Memory & Cognition, 46*(3), 370–383. https://doi.org/10.3758/s13421-017-0771-7

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Kun, A. (1977). Development of the magnitude-covariation and compensation schemata in ability and effort attributions of performance. *Child Development, 48*(3), 862. https://doi.org/10.2307/1128335

Lee, H., Yu, S. L., Lin, T.-J., & Kim, M. (2025). "Am I trying hard or harder than others?": Gender differences in reciprocal relations between perceived effort, science self-concept, and achievement in chemistry. *Contemporary Educational Psychology, 80*, Article 102349. https://doi.org/10.1016/j.cedpsych.2025.102349

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (2023). Extending the reciprocal effects model of math self-concept and achievement: Long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology, 115*(2), 193–211. https://doi.org/10.1037/edu0000750

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction, 35*, 16–32. https://doi.org/10.1016/j.learninstruc.2014.08.005

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology, 52*(8), 1273–1290. https://doi.org/10.1037/dev0000146

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of

lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331–353. https://doi.org/10.1037/edu0000281
Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review, 120*(3), 544–560. https://doi.org/10.1037/a0032459
Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education, 21*(4), 467–487. https://doi.org/10.1007/BF03173515
Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research, 90*(3), 376–419. https://doi.org/10.3102/0034654320919354
Muenks, K., & Miele, D. B. (2017). Students' thinking about effort and ability: The role of developmental, contextual, and individual difference factors. *Review of Educational Research, 87*(4), 707–735. https://doi.org/10.3102/0034654316689328
Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49*(3), 800. https://doi.org/10.2307/1128250
Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*(3), 328–346. https://doi.org/10.1037/0033-295X.91.3.328
Nicholls, J. G., & Miller, A. T. (1984). Reasoning about the ability of self and others: A developmental study. *Child Development, 55*(6), 1990. https://doi.org/10.2307/1129774
Pfeuffer, J., Stoeger, H., & Ziegler, A. (2025). Amplified social inheritance in STEM: Parental careers shape adolescent identity and choice intentions. *International Journal of Adolescence and Youth, 30*(1), Article 2551107. https://doi.org/10.1080/02673843.2025.2551107.
Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology, 84*(1), 152–174. https://doi.org/10.1111/bjep.12028
R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Rhemtulla, M., van Bork, R., & Borsboom, D. (2019). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods, 25*(1), 30–45. https://doi.org/10.1037/met0000220
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. URL http://www.jstatsoft.org/v48/i02/.
Schroeders, U., & Jansen, M. (2022). Science self-concept – More than the sum of its parts? *The Journal of Experimental Education, 90*(2), 435–451. https://doi.org/10.1080/00220973.2020.1740967
Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of reference for academic self-concept. *Educational Psychologist, 37*(4), 233–244. https://doi.org/10.1207/S15326985EP3704_3
Stoeger, H., Beer, A. L., & Ziegler, A. (2025). Students' associations with the STEM acronym and their impact on value beliefs and STEM choices. *Annals of the New York Academy of Sciences, 1552*(1), 297–313.
Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p < 0.05.". *The American Statistician, 73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913
Wilson, A. E., & Ross, M. (2000). The frequency of temporal-self and social comparisons in people's personal appraisals. *Journal of Personality and Social Psychology, 78*(5), 928–942. https://doi.org/10.1037/0022-3514.78.5.928
Wolff, F., Helm, F., & Möller, J. (2019). Integrating the 2I/E model into dimensional comparison theory: Towards a comprehensive comparison theory of academic self-concept formation. *Learning and Instruction, 62*, 64–75. https://doi.org/10.1016/j.learninstruc.2019.05.007
Wolff, F., Nagy, G., Retelsdorf, J., Helm, F., Köller, O., & Möller, J. (2019). The 2I/E model: Integrating temporal comparisons into the internal/external frame of reference model. *Journal of Educational Psychology, 111*(7), 1131–1161. https://doi.org/10.1037/edu0000319
Wolff, F., Sticca, F., Niepel, C., Götz, T., Van Damme, J., & Möller, J. (2021). The reciprocal 2I/E model: An investigation of mutual relations between achievement and self-concept levels and changes in the math and verbal domain across three countries. *Journal of Educational Psychology, 113*(8), 1529–1549. https://doi.org/10.1037/edu0000632

Tobias Debatin is a postdoctoral researcher at the University of Regensburg. His major research interests include competence beliefs, career choices in STEM, learning resources, cognitive ability, and research methods.

Heidrun Stoeger is a full professor at the University of Regensburg. Her primary research interests include mentoring, gender studies in STEM, learning resources, talent development, and learning strategies.

Albert Ziegler is a full professor at the University of Erlangen–Nuremberg. His main research interests encompass mentoring, learning resources, talent development, expertise, and gender studies in STEM.