**On the interplay of motivational characteristics and academic achievement:   
The role of Need for Cognition**

AERA-22-0140

# Revision letter

# Editor

**EC1:** In reading the reviewers' comments, all three indicate that this manuscript has promise for making a substantial contribution to the field. However, the reviewers also had questions about the framing of the study's purpose and theoretical motivation, as well as some of your methodological choices. I share the reviewers' queries about using a sum score approach for the NFC construct, and at a minimum, some justification of your scoring approach would be useful. Both Reviewers 1 and 3 (point 2) raise this issue. Reviewer 2 also raised questions about the nesting structure of the data, sample characteristics, and potentially whether missing data for the physics and chemistry could introduce sample selection bias into your analysis. Finally, Reviewer 2 and 3 provide some useful comments about how NFC is conceptualized with other motivational constructs, and Reviewer 3 highlights contributions of the manuscript that could be further emphasized.

**AR1:** We thank the editor and the three reviewers for their overall positive assessment of our manuscript and for the helpful comments. Below, we provide a point-by-point response to all the comments and an outline of changes made to the manuscript. We hope that we could adequately address all issues raised.

# Reviewer #1

**RC1.0:** Thank you for the opportunity to review the manuscript titled “On the interplay of motivational characteristics and academic achievement: The role of Need for Cognition”. The study examined the predictive relation between NFC and academic achievement by fitting latent change score models to a collection of observed repeated measures. The article is easy to read and follow. The research questions and hypotheses are clearly laid out. Overall, I believe this study could potentially make a positive contribution to the field, with some further clarifications and revisions. But as I’m not an expert on this substantive topic, my questions and comments would be mostly about the research methods:

**RC1.1:** How were the composite measures formed (in terms of ASC, INT, HFS, FOF, and NFC)? Were they computed as the mean score across items? Were they computed as the sum score? This was not clearly stated in the manuscript.

**RC1.2:** I assume no measurement models were fitted to the observed item responses. By computing the composite scores, it is assumed that those items were measured without measurement error. Would this be a reasonable assumption? As the study already fit the models within the SEM framework, why not using measurement models instead of composite scores?

**AR1.1 & 1.2:** In the original manuscript, we used mean scores as composite measures. Due to this and the other reviewers’ comments, we now use factor scores derived from fitted measurement models for the constructs in question. We therefore now state in the manuscript (p. 13):

“We fitted four measurement models of 1) school grades estimated directly from the manifest school grades in order to have the same level of abstraction and to handle missing values for these variables as well, 2) the motivational traits in question, i.e., NFC, Hope for Success and Fear of Failure as well as domain-general and domain-specific 3) ability self-concepts and 4) interests. Separate models were fitted because an analysis of all constructs specified in one model failed to converge. Item-based measurement models were specified except for NFC, where a parcelling approach (Little et al., 2002) was used.”

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 151-173. https://doi.org/10.1207/ S15328007SEM0902\_1

**RC1.3:** The model diagrams in Figure 1 are very confusing. If the T2 variables are omitted (as well as all the relevant paths), then it needs to better clarify in the text and figure notes.

**AR1.3:** We agree that in principle it would be more appropriate to plot the full trivariate LCSMs. Yet, we tried multiple options to convey the important information (i.e., self-feedback and cross-domain coupling together with correlations at T1 and correlated change), but finally decided that we cannot find a way to create a figure that is both comprehensive and comprehensible. Because the paths involving the T2 variables are fixed to one anyway, we therefore opted for omitting these variables and paths. We explicitly state this in the text and the figure legend: “please note that for reasons of simplicity, we omitted to plot the T2 variables throughout panels (B) to (F), because all paths involving these variables are fixed to one.”

**RC1.4:** In Figure 1, “dotted = loadings fixed to zero”? Shouldn’t those paths be fixed to 1?

**AR1.4:** We apologize for this mistake, it now reads “dotted = loadings fixed to one”.

**RC1.5:** P.11 lines 34-36, I appreciate that the study conducted the a priori power analysis. But I don’t understand why the study chose to evaluate the power for testing bivariate correlations, instead of the focal model parameters that are more relevant to their research questions (e.g., path coefficients)?

AR1.5: While we continue to report the simple power calculation based on correlations because it conveys the power of our study in an intuitive way, we also report post hoc power analyses for the latent change score modeling approach used in the *Statistical Analyses* section: “We determined post hoc power via the *semPower.postHoc()* function of the *semPower* package (Moshagen & Erdfelder, 2016) using the following parameters: The latent change score models we tested included all the variables in the study and all possible paths and, thus, was a saturated one with zero degrees of freedom. We tested it against a model where all paths related to NFC (except those that define the latent NFC change score) were fixed to zero, i.e., cross-domain coupling paths, correlations at T1 or correlated change. This model had 22 degrees of freedom. Using this number together with an assumed difference in RMSEA between these two models of .06 and a sample size of *N* = 277, we had a post hoc power of 1-𝛽 = .80 at 𝛼 = .05.”

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling, 23*, 54-60. <https://doi.org/10.1080/10705511.2014.950896>

**RC1.6:** Cronbach’s alpha is not sufficient as the reliability measure as it has strong assumptions of measurement model structure (tau equivalence; also see McNeish, 2018). The best approach, of course, is to build measurement models into the path models. Or at least the study should report other reliability measures that are less restricted in terms of the measurement model structure (e.g., McDonald’s omega).

McNeish, D. (2018). Thanks coefficient alpha, we’ll take it from here. *Psychological Methods, 23*(3), 412.

**AR1.6:** We now report McDonald’s omega along with Cronbach’s alpha, see Supplementary Table S1. It is worth noting that it actually doesn’t make any difference which coefficient for reliability is computed.

**RC1.7:** Regarding the descriptive statistics. If the change scores are of focal interest, the corresponding descriptive statistics (univariate statistics and bivariate correlations) need to be reported for the observed difference scores.

**RC1.7:** Because we now use factor scores derived from measurement models of the variables in question (see above), we do no longer provide descriptive statistics, because – having a mean on zero and an arbitrary standard deviation – they would not be intuitively comprehensible. Instead, we provide tables that contain the correlations of the variables in question at the first measurement occasion (T1), the regression of the change scores on the T1 variables, and correlated change i.e., the correlations of the change scores (see Tables 1-3 as well as Tables S2-S9).

**RC1.8:** P.14 last paragraph, I’m not clear about why the predictors in latent change score model were pre-selected based on their statistical significance in the multiple regression model predicting the T2 grades. First of all, dropping predictors entails very strong theoretical assumptions (the variables are entirely independent) that need to be better justified than merely statistical non-significance. To put in a different way, the structural model should be justified by theory to begin with, and any modification should be backed up by theory, rather than mere statistical significance, especially if the dropped variables are theoretically important confounders. Furthermore, the multiple regression models and the latent change scores are not equivalent. Why not directly testing those path coefficients within the latent change score model?

**AR1.8:** We originally aimed at sparse latent change score models that contain only those variables are relevant for the prediction of school grades, but you are right. We now skip the multiple regression part and include all variables of interest in the latent change score models. The results of these analyses are reported in the main manuscript for overall grades (i.e., GPA) and in the supplement for the specific subjects. Yet, for the figures (Fig1B-F), we maintain to focus on the three main drivers of change in grades, i.e., T1 grades, ability self-concept, and NFC.

**RC1.9:** P. 15 line 7-8: what test was conducted for MCAR? What does ≥ .169 correspond to?

**AR1.9:** We used Little’s test for MCAR as implemented in the R package *naniar*. The *p* ≥ .169 corresponds to the lowest *p*-value of the five tests for MCAR performed, i.e., one test per five school subjects (GPA, German, Math, Physics, and Chemistry). In the revised manuscript, it ow reads: “… missing data in all five variable sets were missing completely at random (MCAR), Little’s tests, .169”.

**RC1.10:** The study reported “correlated change” and its statistical significance. The interpretation is quite vague though. Is it the unconditional correlation between latent change scores? Is it the correlation between the residuals of latent change after controlling for the predictors in the model? The manuscript should be very specific and make the correct interpretation accordingly.

**AR1.10:** We are sorry that we did not specify this issue. We now write (on p. 15, 1st paragraph): “In addition, *correlated change* of the variables of interest can be examined, i.e., to what extent does the change in one variable correlate with the change in another variable after taking into account self-feedback and cross-domain coupling (i.e., to what extent do the residuals of the change scores correlate)”

**RC1.11:** P. 18. Lines 38-40, the statement “we examined mutual influences of change in these variables” is not accurate. The same reason as above. Even if it was the unconditional correlation between the latent change scores that was reported, the correlation/covariance cannot be interpreted as mutual influence.

**AR1.11:** In addition to stating what exactly is meant when using the term “correlated change”, we now avoid to use phrase “mutual influence”.

# Reviewer #2

**RC2.0:** This study investigates longitudinal links among academic achievement (assessed as GPA and grades in four subjects), need for cognition (NFC), and the ability self-concept (ASC) across two measurement time points using trivariate LCS models. The longitudinal findings are largely consistent with previous cross-sectional results and can be used to contribute to our understanding of interrelations among key change processes in education. In addition, a notable strength of the study is the pursuit of a reproducible research approach. However, I found the rationale for selecting these three motivational constructs together with NFC and the modeling approach (LCS models) not well-grounded in prior literature or theory.

I will elaborate on my comments in more detail in the following:

**RC2.1:** In the introduction, the foundation of the choice of these three motivational constructs should be made more evident. As the authors stated, self-concept has been shown to be one of the best predictors of academic achievement. That aside, multiple other motivational constructs are theoretically and empirically linked to academic achievement and could have been chosen (p. 5, ll. 5 - 15). Therefore, the value of the motivational constructs a) interest and b) hope for success/fear of failure should be elaborated.

**AR2.1:** We thank the reviewer for pointing this out. We tried to address this issue together with a comment by reviewer 3 (RC3.8) and explained in more detail the approach we based our selection of variables on. Mainly referring to expectancy-value models, the included variables can be seen as essential ones for the prediction of achievement via motivational aspects. Though there are, of course, many other relevant motivational variables one could include, these are all based on the expectancy-value approach and have a trait-like character which allows for a better comparison when including NFC as additional trait.

The new part reads as follows: “These constructs are part of prominent motivational theories, especially in the context of expectancy-value theories (cf., Atkinson, 1957; Eccles & Wigfield, 2020; Wigfield & Eccles, 2000). As early as in 1957, Atkinson introduced an expectancy-value model based on Murray’s (1938) work (Atkinson, 1957) that comprised essential achievement motives, namely approaching success and avoiding failure, as basis for expectancies for success. Atkinson viewed these motives as relatively stable dispositions describing individual differences in the relative strength of approach and avoidance behaviors, respectively (for an overview see Wigfield et al., 2009). Then in turn, trait-like motivational variables as hope for success and fear of failure can be seen as antecedents for approach and avoidance performance goals, respectively (Elliot & Church, 1997). Based on Atkinson (1957), the expectancy-value theory of Eccles and Wigfield (e.g., Wigfield & Eccles, 2002; Eccles & Wigfield, 2020) comprises the most relevant predictors of achievement motivation and the resulting performance as well as variables influencing these predictors (e.g., cultural or social influences). In this model, expectations of success and values are directly associated with achievement. However, again directly influencing expectations of success and values associated with a task, goals and self-schemata can be find in the model, with the ability self-concept being one of these variables. Ability self-concept, in turn, has proven to be of utmost importance in educational contexts (see below). So, based on the described expectancy-value approach, to get a comprehensive picture of achievement motivation in school, the aforementioned variables should be included. They are each briefly described below.”

**RC2.2:** Furthermore, the introduction does not present a clear and coherent argument supporting the focus on the achievement in physics and chemistry (besides L1 and math).

**AR2.2:** We agree with the reviewer that of course other subjects could also be of interest in this context. The present study examined a convenience sample using data from a larger project on gender differences in STEM subjects, so Grades from Physics and Chemistry were available. These were included in addition to the commonly used grades, that is, Grade Point Average, German, and Mathematics, in order to allow for a more comprehensive picture than previous studies have done in this context. However, since in-depth understanding of the content taught is particularly important in science subjects, these are also subjects for which clarification of the role of Need for Cognition is of particular interest.

We added a short note on the project in the participants information and a rationale for the subjects within the "present study” section.

p. 11: “Both are subjects where an in-depth understanding of the content and models is essential to be able to successfully manage the tasks within the courses in school, so the role of NFC is of special interest in such subjects.“

p. 13: “Data collection took place within a larger project about gender differences in STEM subjects.”

I also have concerns about the following methodological issues and questions:

**RC2.3: The authors used trivariate LCS models to examine longitudinal couplings between achievement, NFC, and ASC. I wonder why the authors are interested in studying interindividual differences in change across time and missed an explanation of the advantages (and limitations) of latent change score models, for example, compared to cross-lagged panel models (Núñez-Regueiro et al., 2022). This lack of an explanation is especially evident after using the results of multiple regression analysis as a selection criterion for the predictors of subsequent achievement.**

**AR2.3: We used LCS models because having only two measurement occasions, we considered it more appropriate than a cross-lagged panel model. Still, we are aware that even for two occasions, alternative modeling approaches exist. LCS modeling simply was an approach one of the authors had employed before for a study involving two time points. We shortly discuss this issue in the Discussion section under Limitations. In the Introduction, however, we think that we already provided a rationale why LCS modeling is a suitable approach in our context. Apart from this, we no longer use multiple regression to select the variables for LCS modeling.**

**RC2.4:** The sample consisted of secondary students from only one school. The sample description should provide more information about the nested data structure (number of classes). Ideally, this nested data structure should be considered using a correction of standard errors. Alternatively, the effects of the nestedness should be estimated and discussed.

**AR2.4:** The sample was collected at two schools. The students attended eleventh grade. In the German school system, there is no longer a fixed class community beginning with eleventh grade; instead, depending on the subject, students come together in different course compositions. The course size comprises an average of 20 students; depending on the course choice of the individual students, there is partial overlap of students, but generally not the exact same composition of students across different courses. Hence, while there is some nestedness, it is very difficult to address. The following information was added on p. 13: “All students attended courses in German and Maths as well as – depending on their course choice – Physics or Chemistry. Course size comprised on average 20 students.“

**RC2.5:** The authors do not detail the reasons for separating students into groups of about 20.

**AR2.5:** Please see our comment above: The size of the courses is the regular course size which was also used for group testing.

**RC2.6:** The rate of missingness is highest for data sets for physics and chemistry. I wonder if these subjects are mandatory for students or if they dropped out (positive selection bias).

**AR2.6:** Please also see our comment above: The number of students in the subjects is different because of the structure of the German school system. While all students have mandatorily to attend courses in German and Math, within STEM subjects there are options for choice as only one STEM subject has to be attended.

**RC2.7:** Changes in grades could be confounded, i.e., by teacher change (when transitioning from 11th to 12th grade) and class composition. Furthermore, changes in grades do not have to reflect changes in performance exactly. After one year of school, even students with lower grades should have acquired more knowledge.

**AR2.7:** We agree with the reviewer that grades are not a truly objective criterion and address this aspect in the discussion section.

“Concerning the prediction of academic achievement it has to be noted that grades are not a truly objective criterion. They do not fully reflect performance, but a whole range of confounding aspects play into it. Course composition can play a role as well as the teachers themselves, teacher changes can bring grade changes, changes in the students' frame of reference can affect motivation and performance alike. As, on the other hand, these aspects enlarge error variance and therewith the risk of not finding associations or influences, respectively, our results represent a relatively conservative estimate of potential associations and predictive values.”

**RC2.8:** I asked myself why the authors did not include additional covariates, such as students’ sex.

**AR2.8:** This decision had their reason in the focus of our study: As potential differences in sex or gender or other sociodemographic aspects were not in the focus of the research questions we refrained from including additional variables in the already very complex models of our calculations.

A few relatively minor issues:

*Abstract*

**RC2.9:** It would be desirable if the authors could mention the number of measurement time points in the abstract and assign the sample sizes to the time points, i.e., by adding an index (NT1 and NT2) (p. 2, l. 18).

**AR2.9:** We now mention the number of measurement time points and give the sample sizes accordingly: “… in a longitudinal approach across two time points in a sample of secondary school students ( = 271, = 255).”

*Introduction/Theory*

**RC2.10:** Please reformulate the last sentence on page 3 (l. 52). The way the sentence is formulated now, it seems as if students’ SC and interest are behaviors associated with success in learning.

**AR2.10:** We changed the sentence. It reads now: “Accordingly, NFC was shown to be related to academic achievement in different stages of academic life (e.g., Ginet & Py, 2000; Grass et al., 2017; Luong et al., 2017; Preckel, 2014; for a meta-analytical review see von Stumm & Ackerman, 2013) and **to motivational variables as well as aspects of information processing** associated with success in learning.”

**RC2.11:** There is an equal sign missing (r = .20) (p. 5, l. 28).

**AR2.11:** Thank you, this was corrected.

**RC2.12:** Please consider including findings of more recent meta-analyses examining the longitudinal relations between achievement and self-concept, such as Guo et al., 2021 or Möller et al., 2020 (p. 5, ll. 26-40).

**AR2.12:** We added information about these meta-analyses in the respective paragraph.

*Statistical analysis*

**RC2.13:** Need for cognition: I guess it should be “seven-point” instead of “four-point” (p. 12, l. 12)

**AR2.13:** Thank you for noticing this mistake, we corrected it.

**RC2.14:** I appreciate the open science approach, but not every software has to be mentioned in the article use (Editor/IDE) R-package knitr, here, ... This information can be given in the appendix.

**AR2.14:** We now mention only the packages used for the main analyses and mention the helper packages in the Supplement (see Supplemental Methods).

**RC2.15:** The variables were separated into *five* sets (p. 13, l. 53)

**AR2.15:** Thank you for noticing this mistake, we corrected it (see also AR3.19).

**RC2.16:** “To impute missing values” (p. 15, l. 10, and l. 52); The FIML algorithm does not impute missing values; only model parameters are estimated. (Enders, 2001, p. 135; Enders, C., & Bandalos, D., 2001)

**AR2.16:** We now use the phrase “handling of missing values using FIML” throughout the manuscript.

*Supplementary tables*

**RC2.17:** The maximum value of GRD2 in table 3 is above 5, whereas the values should range from 0 to 5.

**AR2.17:** Sorry, we made a mistake here: Actually, the range of grades in the respective German state (Baden-Württemberg) is not 1 to 6 but 0.75 to 6. Hence, the figures in the tables are correct, and we now correctly state the range of the grades variables in the Methods section.

Enders, C. K. (2001, January). A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data. Structural Equation Modeling: A Multidisciplinary Journal, 8(1), 128–141. https://doi.org/10.1207/s15328007sem0801\_7

Enders, C., & Bandalos, D. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal, 8(3), 430–457.  
https://doi.org/10.1207/S15328007SEM0803\_5

Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021, March 23). A Meta-analysis of the Longitudinal Relationship Between Academic Self-Concept and Academic Achievement. Educational Psychology Review, 33(4), 1749–1778. https://doi.org/10.1007/s10648-021-09600-1

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020, April 30). A Meta-Analysis of Relations Between Achievement and Self-Concept. Review of Educational Research, 90(3), 376–419. https://doi.org/10.3102/0034654320919354

Núñez-Regueiro, F., Juhel, J., Bressoux, P., & Nurra, C. (2022, July). Identifying reciprocities in school motivation research: A review of issues and solutions associated with cross-lagged effects models. Journal of Educational Psychology, 114(5), 945–965. https://doi.org/10.1037/edu0000700

# Reviewer #3

**RC3.0:** The manuscript AERA-22-0140 entitled “On the Interplay of motivational characteristics and academic achievement: The Role of Need for Cognition” investigates whether the construct NFC explains incremental variance in academic achievement across different school subjects over and above other motivational constructs (e.g., academic self-concept) and prior academic achievement in a sample of secondary school students. By applying latent change score models in a longitudinal study design, the authors identified NFC to incrementally predict academic achievement as reflected in GPA, German, and Physics grades over and above prior achievement and domain-general or domain-specific academic self-concept, respectively. Based on their findings, the authors propose to include NFC more systematically in models exploring academic achievement in the educational setting.

The present manuscript is addressing a relevant topic and by including prior academic achievement in a longitudinal study design, it aims at generating new evidence on whether NFC is incrementally predicting academic achievement and it is therefore likely to appeal to a broad scientific audience that is interested in constructs that can contribute to a comprehensive explanation of achievement in an educational setting. The manuscript is well structured and the introduction leads logically up to the research questions and hypotheses. Commendably, the authors have shared their data and a comprehensive script that allows to understand and to replicate the different steps of their analyses.

While making a valuable contribution to the field of educational research, a number of aspects the authors could improve upon in order to further strengthen the quality of their manuscript has been identified and will be described in more detail in the following. Once these aspects have been taken into consideration, the manuscript should be considered for publication.

*Major aspects*

**RC3.1:** Clearer definition of the research gap – In the current version of the manuscript, it is rather difficult to understand which new insights the study is offering. In the section The present study, the authors state the following: “Furthermore, up to now, prior achievement was not integrated as performance predictor in studies examining NFC.” If understanding the study by Lavrijsen et al. (2021) cited by the authors correctly, Model 8 of their analyses (p. 781) does however include a control for prior achievement, and they conclude that motivational constructs [including NFC] were identified as significant predictors of academic achievement despite controlling for prior achievement. In this context, the authors of the present study should underline that their own methodological design (e.g., assessment of all measures at two time points as compared to Lavrijsen et al.; see Table 1 on p. 777) displays important statistical advantages. It are those advantages that allow to generate more solid insights when controlling for prior achievement while including the measure in itself has previously been done in research. In addition, the authors should stress more clearly that their study does generate new knowledge on positive reciprocal relations between NFC and academic achievement – two variables that appear to mutually strengthen or weaken each other as stated by the authors in their manuscript (p. 21).

**AR3.1:** We thank the reviewer for their positive evaluation of the value of our study. We rewrote the paragraph at the end of our introduction and pointed out more clearly in which ways we go beyond existing research.

The new part reads as follows: “We follow-up on cross-sectional (e.g., Keller et al., 2019; Luong et al., 2017), and the few longitudinal studies (Preckel, 2014; Lavrijsen et al., 2021) on the role of NFC in predicting academic achievement that examined NFC together with established motivational characteristics. By addressing more school subjects than usually examined, considering prior achievement, and assessing all variables at two points of time in a sample of secondary school students, we go beyond previous work to provide new insights in the interplay of academic achievement, NFC and motivational variables and the incremental value of NFC in this context.”

Additionally, we adapted the part “The present study” to state more clearly the value and novelty of our study:

“Overall, NFC has been proven to be a promising predictor of academic achievement over and above other motivational constructs (e.g., Ginet & Py, 2000; Keller et al., 2019). Yet, so far the evidence on its incremental predictive value is limited by the mainly cross-sectional nature of available studies (e.g., Luong et al., 2017) and by the fact that only a few school subjects (mostly GPA, German, and Math) were considered (e.g., Preckel, 2014). Furthermore, up to now, prior achievement was only integrated as performance predictor in studies examining NFC in one study (Lavrijsen et al, 2021). This is a limitation insofar as besides students’ cognitive abilities their prior achievement is a relevant predictor of future academic achievement (e.g., Hailikari et al., 2007; Steinmayr et al., 2019).

With the present study, we aim at adding to the existing body of research by examining NFC, well-established trait-like motivational indicators routed in expectancy-value approaches (ability self-concept, hope for success and fear of failure, interests, each of them general and subject-specific) and academic achievement (assessed via GPA, and grades in German, Math, Physics, and Chemistry) each at two points of time. In doing so, we will be able to extend insights by Lavrijsen et al. (2021) who did assess NFC only at one point of time. Furthermore, by considering GPA plus four subject grades, we extend the existing literature on predicting academic achievement in school not only in general and in the domains of Math and German (see Steinmayr & Spinath, 2009), but also on focusing on the further domains Physics and Chemistry. Both are subjects where an in-depth understanding of the content and models is essential to able to successfully manage the tasks within the courses in school, so the role of NFC is of special interest in such subjects. By applying latent change score modelling, we will be able to determine the influence of our different predictors on the change of academic achievement in general and in different domains in school over time. At the same time, mutual influences of changes in academic achievement, NFC and motivational constructs can be detected (i.e., correlated change). As it is well-known that there are reciprocal relations between academic achievement and ability self-concept (see Guay et al., 2003; Wu et al, 2021) it is of special interest to examine such potential relations for NFC as well.”

**RC3.2:** Usage of latent variables for each scale instead of sum scores – In their study, the authors applied latent change score modelling in order to explore their research questions. Using latent variables has some important statistical advantages (e.g., Little, 2013) over manifest variables and in this context, it seemed surprising that the authors relied on the creation of sum scores for each of their scales before including them as manifest variables into their models (as displayed in Figure 1). In the reference by Kivet et al., (2017) that has been cited by the authors, the following has been stated: “We can easily extend this model to have an explicit measurement model by replacing the observed score with a latent variable, measured by a set of observed variables. We refer to this representation as a multiple indicator latent change score model, as our aim is to model change in the latent score rather than observed scores. […] A further extension of the latent change score model is to include a second (or third, fourth, etc.) domain of interest. For convenience in notation and graphical representation we will revert back to using only observed scores, but all extensions can and – where possible should – be modelled using latent (multiple indicator) factors”(p. 4). In order to make use of the full statistical advantages of latent change score modelling, the authors should therefore include latent variables into their model that represent each scale. In case this is not possible (e.g., non-identification of the model), the authors should clearly state the reasons for using manifest indicators in their methods section and discuss potential implications in the limitations. Also, if the authors keep using manifest variables, please specify if missing values were imputed after the creation of sum scores or previously at item level.

**AR3.2:** As noted above (AR1.1-2), we now use factor scores derived from measurement models of the variables in question and provide information on this procedure on p. 13.

Minor aspects

Besides these two major aspects, the authors should address a number of minor aspects in order to improve the manuscript’s flow and readability:

*Introduction*

**RC3.3:** The formulation “Over the last years” (p. 3) reads rather short-termed for the actual duration of research focusing on the relation between NFC and academic achievement (e.g., Cacioppo & Petty, 1982; Ginet & Py, 2000; Tolentino et al., 1990).

**AR3.3:** Thank you for pointing this out, we reformulated the sentence. It reads now: “Initially introduced in the context of social psychology, increasingly, an additional predictor of academic achievement came into the focus of research in this field: the personality trait Need for Cognition (NFC), defined as the stable intrinsic motivation of an individual to engage in and enjoy challenging intellectual activity (Cacioppo & Petty, 1982; Cacioppo et al., 1996).”

**RC3.4:** When defining NFC as “the stable intrinsic motivation of an individual to engage in and enjoy challenging intellectual activity” (e.g., p. 3 and 6), please refer to the original reference (Cacioppo & Petty, 1982), in which the construct of NFC has been introduced, instead of citing a later one (Cacioppo et al., 1996).

**AR3.4:** Please see the sentence above. We added the original reference and cite both, Cacioppo and Petty (1982) and Cacioppo et al. (1996), to give the readers the information about the first relevant publication but also the source of the most frequently used definition that can be found in Cacioppo et al. (1996).

**RC3.5:** Please clarify what is meant by “a large longitudinal study” (p. 4). Was the study large in terms of its N or did it include many time points?

**AR3.5:** We clarified the information by writing: “Only last year, a large longitudinal study **with over 3.000 Flemish Grade 7 students** examined **a comprehensive set of variables** including intelligence, the Big Five, a range of different motivational measures, and NFC in order to determine their value in predicting academic achievement in school (Lavrijsen et al., 2021).“

**RC3.6:** The first part of the introduction contains some information that does not seem necessary or that is repeated at other places of the manuscript, where it fits better. On page 3 for example, the authors define investment traits without however referring back to such traits in the remaining manuscript. As it rather seems to break the flow within the paragraph focusing on NFC, it may be deleted. The sentence “Concepts like ability self-concept, hope for success and fear of failure, interest and values are well known and equally established indicators (Wigfield & Cambria, 2010; e.g., Wigfield & Eccles, 2000) that are subsumed under the umbrella term of achievement motivation (Steinmayr et al., 2019)” is for example repeated in similar wording in the section Achievement Motivation and its relation to academic achievement, where it seems more relevant. For the sake of the introduction, it might be enough to include examples of motivational variables in brackets of the preceding sentence.

**AR3.6:** We shortened the paragraph about achievement motivation as suggested and put the information about investment traits at another part of the manuscript to illustrate the difference between typical and maximum cognitive performance.

**RC3.7:** Along the lines of the first major aspects addressed further above, the last paragraph before the section Achievement Motivation and its relation to academic achievement should be adapted in order to express more clearly which previous research findings the present study is aiming at strengthening (e.g., Keller et al., 2019, Lavrijsen et al., 2021) and which entirely new insights it is providing through its longitudinal design.

**AR3.7:** Please see our comment AR3.1. – we rewrote this paragraph.

*Achievement Motivation and its relation to academic achievement*

**RC3.8:** In this section, the authors refer to “prominent motivational theories” (p. 5) and the “influential model of Wigfield and Eccles” (p. 6). For readers that are not familiar with these theories/models, please consider explaining them in a few sentences or reflect whether referring to them is actually contributing meaningfully to the present manuscript; otherwise, these references regarding specific models and/or theories may be excluded while only stating the references that have introduced the constructs in question.

**AR3.8:** Thank you for pointing this out. We briefly explain the models now in order to give readers a bit more background and to more convincingly derive the choice of our variables (which was a point raised by reviewer 2., RC2.1).

This is the new paragraph: “These constructs are part of prominent motivational theories, especially in the context of expectancy-value theories (cf., Atkinson, 1957; Eccles & Wigfield, 2020; Wigfield & Eccles, 2000). As early as in 1957, Atkinson introduced an expectancy-value model based on Murray’s (1938) work (Atkinson, 1957) that comprised essential achievement motives, namely approaching success and avoiding failure, as basis for expectancies for success. Atkinson viewed these motives as relatively stable dispositions describing individual differences in the relative strength of approach and avoidance behaviors, respectively (for an overview see Wigfield et al., 2009). Then in turn, trait-like motivational variables as hope for success and fear of failure can be seen as antecedents for approach and avoidance performance goals, respectively (Elliot & Church, 1997). Based on Atkinson (1957), the expectancy-value theory of Eccles and Wigfield (e.g., Wigfield & Eccles, 2002; Eccles & Wigfield, 2020) comprises the most relevant predictors of achievement motivation and the resulting performance as well as variables influencing these predictors (e.g., cultural or social influences). In this model, expectations of success and values are directly associated with achievement. However, again directly influencing expectations of success and values associated with a task, goals and self-schemata can be find in the model, with the ability self-concept being one of these variables. Ability self-concept, in turn, has proven to be of utmost importance in educational contexts (see below). So, based on the described expectancy-value approach, to get a comprehensive picture of achievement motivation in school, the aforementioned variables should be included. They are each briefly described below.”

**RC3.9:** On p. 5, the authors finish a sentence by stating “which is why they were included in this study as important motivational indicators”. Such a statement does not belong in the theoretical overview but rather into the section The present study.

**AR3.9:** We deleted this sentence.

**RC3.10:** To increase readability, please include “see also Eccles and Wigfield (2020)” into the preceding brackets (p. 6).

**AR3.10:** We changed the sentence accordingly.

*Need for Cognition and academic achievement*

**RC3.11:** On page 7, the authors write the following sentence: “NFC, mirroring the typical cognitive performance of a person, has been shown to be rather modestly related to intelligence and its fluid (Fleischhauer et al., 2010) and crystallized (von Stumm & Ackerman, 2013), components”. It does not become clear what exactly the authors mean by “mirroring the typical cognitive performance of a person”. Do they want to state that NFC as a construct mirrors typical cognitive performance of individuals or rather that findings regarding the relation between NFC and intelligence mirror those between typical cognitive performance and intelligence? Please clarify this sentence and do also add a reference, in which this has been previously found.

**AR3.11:** We rewrote the paragraph to illustrate this aspect more clearly. It reads now: “Conceptually, NFC belongs to the group of investment traits (von Stumm & Ackerman, 2013). These traits determine how individuals typically invest their cognitive resources and how they deal with cognitively challenging material. As such, NFC mirrors the *typical* cognitive performance of a person while intelligence as an ability trait represents the potential *maximum* cognitive performance (von Stumm et al., 2011).“

**RC3.12:** As the paragraph on how NFC is related with academic achievement across different stages of school seems particularly important for the theoretical background of the present study, it would be beneficial for an even deeper understanding of previous research findings if the authors could include details on the different stages of school (e.g., which grades – primary or secondary school – were analyzed in which respective study). In addition and especially in light of the broad subjects included in the present study, it would be worth to specifically state which subjects were included in previous work (e.g., Ginet and Py, 2000; Colling et al., 2022). If these aspects are introduced in the theoretical background of the study, the authors can more easily extend their discussion in this direction as suggested further below.

**AR3.12:** We added information about the respective studies.

The paragraph reads now: “NFC correlates with academic achievement across different stages of school and university: For example, in a longitudinal study, examining over 700 secondary-school students (grade 5 at T1), Preckel (2014) found a weak positive correlation primarily for Math in secondary school. NFC incrementally predicted grades in math over and above intelligence at T2 and T3. Ginet and Py (2000) found a mean correlation of between NFC and academic achievement (average from grades in French, math, and English) in school across all school years studied, with lower correlations (*r* =.10, N = 50) in earlier and higher correlations (*r* = .50/.42, N = 39/50) in later school years, a pattern that can also be found in Luong et al. (2017). While there were practically no associations in grade 3, associations were about *r* = .30 in grade 6, and 9, respectively, in a large sample of over 4.000 Finnish students. Examining over 3.000 Luxembourg students in 9th grade, Colling et al. (2022) also report differences in the strength of the correlations with academic achievement in school, here depending on the type of school, with the associations between NFC and academic achievement being strongest in the highest and weakest in the lowest school track.“

*NFC and motivational aspects of learning*

**RC3.13:** The section on NFC and motivational aspects of learning seems slightly unbalanced with regard to the study’s main research interest. The first two thirds of the section focus on correlations between NFC and different motivational aspects of learning (e.g., deeper processing, academic self-concept) while a considerably shorter part of the section reports findings from studies that analyzed whether NFC explains incremental variance in achievement over and above other motivational constructs. Considering the main research of the present study, the authors should considerably shorten the first part of the section and extend on the second part by reporting the findings of studies that are more closely related to their own research interest in more detail (e.g., Keller et al., 2019, Luong et al., 2017, Meier et al., 2014, Lavrijsen et al., 2021). In order to embed their own findings more strongly into previous research in the scope of the discussion, details on the considered constructs, on the respective samples and educational settings seem worth to be shared. Also, results from the longitudinal study by Preckel (2014) finding NFC to explain incremental variance over and above intelligence could be referred to in this context.

**AR3.13:** We thank the reviewer for pointing out the imbalance of the depiction of the state of research. We added the suggested information about the studies that examined the incremental value of NFC to give the readers a more comprehensive picture about existing studies. However, we refrained from shortening the first paragraph of this section in order to allow for detailed information about associations of NFC to relevant motivational aspects of learning.

Several studies examined NFC along with other motivational variables and found NFC to explain variance in academic achievement beyond established motivational variables such as learning orientation or ability self-concept (Keller et al., 2019; Luong et al., 2017). As mentioned above, Preckel (2014) demonstrated incremental validity of NFC over and above intelligence in the prediction of math achievement in a sample of grade 5 students. Keller et al. (2019) examined the incremental validity of NFC in the prediction of academic achievement in three samples from Luxembourg (grade 9), Finland (grades 6 and 9) and Germany (grades 3 and 4). NFC incrementally predicted performance in Math and German or Finnish, respectively over and above ability self-concept and interest in the Finnish and Luxembourgish sample and – to a smaller amount – in German in the 4th grade of the German sample. Luong et al. examined the relevance of NFC in a Finnish sample of over 4.000 students (from the 3rd, 6th and 9th grade; 10 to 16 years of age). In the overall sample and in school years 6 and 9, NFC was a significant predictor of academic achievement along with ability self-concept, control motivation, and learning orientation. Meier et al. (2014) examined potential predictors of the attendance of a gifted class in a sample of about 900 students attending grade 5. They found that NFC, compared to other motivational constructs like academic interests and goal orientations, significantly predicted the attendance of a gifted class even when controlling for cognitive ability and other factors like parental education level or ability self-concept. Lavrijsen et al. (2021) longitudinally examined the predictive value of intelligence, personality (Big Five and NFC) and different motivational constructs (e.g., autonomous/controlled motivation, achievement motives and goals) for academic achievement in a sample of 3.409 Flemish Grade 7 students. They found intelligence, NFC, and the ability self-concept to be the strongest predictors of Math grades and performance in standardized Math tests.

*The present study*

**RC3.14:** Along the lines of the first major aspect addressed further above, please adapt the first paragraph in order to express more clearly which previous research findings the present study wants to strengthen (e.g., Keller et al., 2019, Lavrijsen et al., 2021) and which entirely new insights it is providing through its longitudinal design (e.g., research question 3).

**AR3.14:** Please see our comment AR3.1. – we rewrote this paragraph.

**RC3.15:** When presenting their research questions and assumptions, it would be helpful for the reader if the authors would include previously introduced references directly behind “Because of evidence” (RQ1) and “Based on previous findings” (RQ2).

**AR3.15:** We added illustrative references.

*Methods*

**RC3.16:** In the section on Participants, it feels like the term women does not fit the age range of students very well. As suggested in the APA 7th edition, the authors could refer to their sample by using the term of female adolescents.

**AR3.16:** Thank you for this suggestion. We reformulated the respective sentence that now reads: “We eventually managed to recruit a sample of = 277 adolescents (60% female) at the first measurement occasion (T1) of which = 251 adolescents (61% female) also took part at the second measurement occasion (T2) …“

**RC3.17:** In the section on Participants, the authors state the following: “Yet, we tried to impute missing values to raise power (see below, Statistical analyses).” The term tried somehow implies that they did not succeed, whereas the section on Statistical analyses describes how missing values seem to have been successfully imputed using FIML. Please specify or adapt the terminology accordingly.

**AR3.17:** We now write in the last sentence of the participants section: “Yet, we used an approach to handle missing values to raise power (see below, *Statistical analyses*).”

**RC3.18:** In the section Material, authors report Cronbach’s αs and retest reliability values for the scales used in their study. Before the screening of Table 1, it has remained unclear whether these values were from the present study or from previous research. Please specify in the respective sentences, for example by adding elements such as “In previous research, the scale showed high internal consistency.”

**AR3.18:** We hope that the revised wording in the Method section now makes clear that we refer to previous research when reporting reliability coefficients, e.g., “The scale has been shown to exhibit comparably high internal consistency, Cronbach’s *α* > .80 (Bless et al., 1994; Fleischhauer et al., 2010), and retest reliability, *rtt*=.83 across 8 to 18 weeks (Fleischhauer et al., 2015)” or “Both scales exhibited high internal consistencies in previous research, Cronbach’s *α* ≥ .85 (Steinmayr & Spinath, 2009)”.

**RC3.19:** In the section Statistical Analysis, the following sentence proved difficult to read: “First, the variables were separated into four sets, each containing the T1 and T2 measurements of the variables Hope for Success (HfS), Fear of Failure (FoF), and Need for Cognition (NFC) as well as either GPA, overall ability self-concept regarding school, and general interest in school, or domain-specific grades, ability self-concept and interest in German, Math, Physics, and Chemistry.” A possibility to increase readability would be to write “as well as either domain-general (name all variables) or domain-specific measures (name all variables)”.

**AR3.19:** Thank you very much. We followed you suggestion and now (also responding to RC2.15) write: “First, the variables were separated into five sets: Each set contained the T1 and T2 measurements of the variables Hope for Success (HfS), Fear of Failure (FoF), and Need for Cognition (NFC) as well as grades, ability self-concept, and interest in school in general as well as in German, Math, Physics, and Chemistry in particular.”

*Results*

**RC3.20:** In order to improve readability in the results section, authors should consider putting their results into brackets. In addition, the authors should reflect on whether it might be enough to only report standardized coefficients and p-values within the text and to refer to the respective tables for further values (e.g., unstandardized coefficients, confidence intervals). The values would then also be in line with those displayed in Figure 1. Similarly, including all model fit values within the text has a negative impact on readability. The authors could create a table for all fit values and only report χ^2 difference test results.

**AR3.20:** We now only report standardized coefficients and associated p-values in the text and provide more detailed statistics in the tables. We now also report only one result of a χ2 difference test.

*Discussion*

**RC3.21:** In line with the previous comment on the potential extension of the theoretical background with regard to which subjects were included in previous works, the authors should relate in more detail how their own findings (a) strengthen or stand in contrast to previous findings (e.g., verbal school subjects, Math) and (b) generate new knowledge on additional subjects (e.g., Chemistry). As some of the subjects show certain similarities (e.g., STEM), the authors could elaborate how this might impact their findings and discuss in the limitation sections how the inclusion of other subjects that differ both in their content and in the applied pedagogical approaches (e.g., History) might broaden the body of knowledge on the relation between NFC and academic achievement.

**AR3.21:** We agree with the reviewer that these findings could be more differentiated described in the discussion. Accordingly, we added additional information and pointed to the necessity to include other subjects in future studies.

“With regard to the four subjects, NFC showed the third-highest correlations with grades, after domain-specific ability self-concept and domain-specific interest. Correlations were medium-sizes for all subjects which is comparable to findings, for example by Ginet and Py (2000) or Luong et al. (2017). While previous findings usually focused on GPA, Math, and first language (e.g., German, French), our findings extend the knowledge of NFC and academic achievement to two STEM subjects, namely Physics and Chemistry. The medium-sized associations (about *r* = 0.35) highlight the importance of NFC in subjects that require conquering the models and approaches to get an in-depth understanding which is an inherent conceptual aspect of NFC (Cacioppo et al., 1996).”

„At the background of the findings concerning Physics and Chemistry as STEM subjects and the potential that was shown for NFC it would be interesting to include a broader range of subjects in future studies to be able to examine differences in the predictive value of NFC in subjects with different characteristics or requirements, respectively.”

**RC3.22:** Based on the suggestion to extend the theoretical background by reporting the findings of studies that are more closely related to their own research interest in more detail (e.g., Keller et al., 2019, Luong et al., 2017, Meier et al., 2014, Lavrijsen et al., 2021), the authors could embed their own findings more strongly into the current state of the art (e.g., studies that included self-concept as a measure).

**AR3.22:** Thank you for this suggestion. We address this point in the discussion section by embedding our findings in the state of research. We do so in different parts of the discussion (please see also AR3.21) and give two illustrative examples in this letter:

“NFC showed positive concurrent and predictive correlations with achievement and all motivational variables, except for fear of failure for which concurrent and predictive correlations were negative. This correlational pattern was found for domain-general measures as well as for the four subjects. Correlations were of medium to large effect size and comparable to previous findings: We found strong associations of NFC with ability self-concept (comparable to, e.g., Dickhäuser & Reinhard, 2019), medium-sized correlations with interest (comparable to Preckel, 2014, or Keller et al., 2019), and a strong positive relation to hope for success and at the same time a medium-sized negative association with fear of failure (see e.g., Diseth & Martinsen, 2003).“

„To conclude, with regard to all grades examined and comparable to the results of Keller et al. (2019), or Lavrijsen et al. (2021), respectively, NFC proved to be a valuable predictor of academic achievement besides prior achievement and ability self-concept. Taking a differentiated view, compared to one of the best established predictors in educational research, the ability self-concept, NFC was even broader able to predict academic achievement. As mentioned above, extending previous findings by including two STEM subjects highlights convincingly that NFC enfolds its potential especially in subjects that require deeper thinking.“

**RC3.23:** In the theoretical background, the authors introduce Hope for Success/Fear of Failure and Task Values/Interest as potential further variables of interest in their study and reported how previous studies identified significant relations between those constructs and academic achievement. After describing in their statistical analysis that these variables did not identify as significant predictors of academic achievement in their sample, they are no longer mentioned within the manuscript. In the discussion, the authors should, however, reflect on why these variables might not have identified as significant predictors. In this context, the items of Hope for Success/Fear of Failure read highly similarly to established NFC items and these variables also displayed medium to high correlations as indicated in the manuscript’s tables. The authors could reflect on whether this observation might have an impact on their finding or discuss potential other explanations shortly as it currently seems as if the argumentation regarding these excluded variables does not come full circle.

**AR3.23:** Thank you for pointing this out. We elaborate now on this issue in the discussion.

„For GPA, and German, prior achievement positively predicted changes in grades, as did NFC and general, or domain specific ability self-concept, respectively. For German, Hope for Success was another relevant predictor but it is noteworthy that it seemed to be largely redundant to NFC in most of the analyses which is also indicated by the high (*r* = .82) intercorrelation of both variables. That the items of both scales had higher similarities might be one reason for this finding. The differentiation of both variables should be a subject of further studies.“

**RC3.24:** In the section Limitations and further directions, the authors state that their sample cannot be considered as representative for the German population of adolescents. Please specify the reasons why in more detail (e.g., only academic track students, potentially higher SES students, data from only one Bundesland).

**AR3.24**: We clarified this by writing: “… it was not representative for the German population of adolescents as we assessed data only in two schools of one German federal state.“

*General comments*

**RC3.25:** As requested by the submission guidelines of AERA OPEN, the authors should adapt their in-text citations to the style guidelines of Publication Manual of the American Psychological Association (APA), 7th edition (e.g., Deary et al., 2007 instead of Deary, Strand, Smith, & Fernandes, 2007).

**AR3.25:** We now adhere to the APA7 citation style throughout the manuscript.

References mentioned that are not a part of the present manuscript

Little, T. D. (2013). Longitudinal structural equation modeling. The Guilford Press.