Environmental Psychology Open

Peer review process documentation

of the article

Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school students' civic engagement for climate protection

by

Amelie Spliesgart, Stephan Heinzel, Phillip Gutberlet, Marie Heitfeld, Paula Blumenschein, Vivian Frick & Jan Keller

published in

Environmental Psychology Open

DOI: https://doi.org/10.69805/epo.v29.a25

Handling editor: Mathias Hofmann

Environmental Psychology Open

Review Round 1

Review A

First, I thank the authors and the editors for giving me the opportunity to read this interesting manuscript. The manuscript presents a validation study for a newly developed scale to measure civic engagement for climate protection among adolescents. In a sample of $N = 360$ secondary school students, the authors find evidence for the unidimensionality of the scale, its high reliability, and the expected relations with theoretically derived psychological constructs and other classes of behavior. I think that this manuscript can make an important contribution based on its focus on an understudied population and an often-overlooked behavioral type in environmental psychology. However, I believe that this manuscript should be improved before publication. Below, I have listed some major and minor concerns:

The abstract gives a nice overview of the study and its findings. But the last statement is too general and should mention the limitation that the scale has only been tested with German students/adolescents so far. Currently, the abstract suggests general applicability. As a minor note, M_{age} is formatted differently in the German and English abstract.

Turning to the main text, I see two major issues: one is theoretical and the other methodological.

Starting with the theoretical issue, civic engagement for climate protection is not clearly enough differentiated from social influence and collective action (some aspects of this are mentioned in the discussion, but it would benefit the reader, if this was moved to Section 1.1). The authors nicely derive how civic engagement is different from individual consumption, but it would be helpful to clearly define the relation between civic engagement and collective action and social influence (in the paragraph that introduces civic engagement). Throughout the manuscript, I struggled with distinguishing between the different types, for example, as "collective efforts" is part of the current definition of civic engagement, and Item 7 of the civic engagement scale and Item 4 of the social influence scale overlap. In addition the authors speak of wanting to understand how people advocate for climate protection in the first sentence of Section 1.2 which would also fit the description of social influence (e.g., Abrahamse & Steg, 2013).

Besides this definitional aspect, I would question what construct the authors measure with the civic engagement scale. The reasoning in the introduction is coherent, however in Section 1.1 it is unclear whether the scale is supposed to measure a person's ecological handprint (person property; see Lange, 2024) or a property of the behavior (e.g., frequency of specific behaviors). I highly recommend reading Lange (2024) to determine what exactly they are measuring and to adjust the definition accordingly. If the authors come to the conclusion that they measure a person property, I would question whether this property (i.e., the intention to protect the climate by "focusing on systemic causes of environmental problems and the promotion of environmental sustainability through collective efforts") is different from the intention to protect the climate by reducing personal consumption or influencing others. Previous work shows that these different behavioral types might have the same underlying intention (e.g., Kaiser, 2021; Kaiser & Lange, 2021). The high correlations between the different behavior self-report scales (even in light of restricted variability and reliability) hint towards this interpretation.

I think the point you are making in the first paragraph on Page 6 is really interesting: that different forms of civic engagement might just represent different levels of a unidimensional construct. This could very well also be the case for different forms of climate-protective behaviors.

When it comes to the methodology, I do not think that Classical Test Theory (CTT) is the most appropriate methodological approach to answer the research questions (especially, as the authors explicitly want to evaluate item characteristics), and I would suggest several additional analyses before publication of this manuscript.
It is not clear to me why the authors used CTT to evaluate item characteristics when Item Response Theory (IRT) would have clear advantages in this regard (e.g., Lalot, 2025). I would strongly suggest to test the construct dimensionality and evaluate the scale using IRT methods like the Partial-Credit model, Graded Response model or dichotomous Rasch model (would necessitate prior item dichotomization). If you stick to CTT, I would expect an explanation why it was chosen over IRT.
Furthermore, the authors are unclear about the evaluation of the correlations for Research Question 4: At what *r*-level are the correlations in line with the assumptions (e.g., is a strong correlation not in line with discriminant validity and a weak correlation not in line with the appropriateness of the nomological network)? In addition, it might be interesting to investigate how the different behavior classes vary in their relations with the psychosocial predictors (if they are not found to represent the same psychological property). For a thorough scale validation and evaluation, I would also expect the following tests:

- Does the scale measure the construct equally across groups (measurement invariance or differential item functioning); for example, I would expect some items to be differently understood by 17-year-olds vs. 12-year-olds, which could compromise the scale fairness?
- Going beyond HTMT, it should be tested whether the three behavior self-report scales (individual consumption, social influence, civic engagement) can be represented in a one-dimensional measurement models (either with EFA, CFA, or preferably a IRT method).
- The authors should evaluate the dimensionality and measurement models of all scales they use as the relation between the civic engagement scale and the other scales is such an integral part of this manuscript.

Also, in case the authors stick to CTT, they should compute CFAs, which would allow to evaluate the model fit based on indices like CFI, RMSEA, and SRMR.

In addition to these two issues, I have the following remarks:
In Section 1.3, it is not clear why the authors chose these two specific theories for the nomological network. This should be made more explicit. Currently, they only state that both individual and collective variables seem to be important. But why did they then specifically choose the NAM and not the Value-Belief-Norm theory (Stern et al., 1999), for example, for the individual factors? I am not saying that the choice of theories is bad, just that it should be made more explicit why they were chosen for this study. Also, the SIMPEA is presented as if the focus of it is on civic engagement for climate action, when this is not the case (Fritsche et al., 2018).
There are no explicit hypotheses in this manuscript which makes the evaluation of the research questions difficult. Hypotheses could be added to improve understandability, for example in Section 1.5.
In general, I was wondering whether difficult terms like petition were explained? Was the understanding of items and introductory text checked? To me the wording seems quite complex for the younger participants of the study (e.g., Verankerung, Verbündete, Petition). Were there any pretests to ensure understandability? In addition, asking a 12-year-old whether they donated money might not be suitable. These aspects of the study should be discussed as they might have affected the findings.

Environmental Psychology Open

Regarding the results, the part-whole correlations were not described in Section 2.4 and there is neither a mention nor a reference of the acceptable range. Furthermore, there is a nice description of how missing values were treated on the civic engagement scale, but there is no mention of how missing data on all other constructs were treated and distributed. This would be important information to present.
I understand the reasoning by the authors that civic engagement behavior is rarely shown and that either a dichotomous scale or frequency scale would have led to restricted variability. But at the same time artificially creating a 6-point Likert-scale for a truly dichotomous variable (behavior shown yes-no) can also affect the reliability of the responses (Kaiser & Wilson, 2000). This could be discussed or mentioned.
When it comes to the sample description it would improve readability to either define social status already in the corresponding section (i.e., what was the item) or refer to the section where it is described. Generally, it would be an important information whether the assessed scales for the variables in the nomological network were previously validated in a similar sample. For example, I would assume that the social ladder scale might be perceived a lot differently by children/adolescents compared to adults (e.g., more volatile).
On page 20, it is written that "Item 7 ("tried to make climate-friendly behavior easier or more approachable in my environment") received significantly higher ratings than the other items." I am wondering whether this significance was tested. If yes, the corresponding test statistic should be presented. The discussion of item 5 is missing that donating is probably the least accessible behavior for children and adolescents which could have the observed negative effect on the correlations.
The discussion of the nomological network should be expanded. Identifying knowledge gaps as an effective point to intervene is a bit counterintuitive given a) the small effects of information interventions (e.g., Bergquist et al., 2023), and b) that subjective knowledge in your study has (descriptively) the weakest correlation with civic engagement.
Finally, I am missing in the discussion a) a CAVEAT for the low reliability of the individual consumption scale (which could be taken into account in the correlation with civic engagement by computing error-attenuation corrected correlations), and b) the possibility that the negative correlation between age and civic engagement might be rooted in younger students not having understood some of the items.

Minor comments:
There is no in-text reference to Figure 1.
Page 6: 1.4 is two times in the heading
Page 8 and 9: The categories for municipality size in the text do not match the categories in the table (100,000-1,000,000 vs 100,000-999,999); in the table the groups are missing a data point: a size of 1,000,000 is not represented.
Page 9: The statistical abbreviations M, SD and n should be in italics.
Page 10 (table note): Please explain why there are missing values / why the numbers not always add up to the total sample size. Are the percentages based on the total sample size or the sample size that provided an answer to the item?
Page 13: The item for knowledge measures subjective knowledge and should be labelled as such. Currently, one might get the impression that a knowledge test was assessed when not reading the measures section.
Page 13: This cannot changed, but the scale for measuring self-efficacy beliefs rather measures beliefs about the ability for social influence and not civic engagement in my interpretation of the items ("Umfeld" as social

environment). This just highlights how important a clear definition and differentiation from related constructs is which should be added in the theoretical background. It would be interesting to compute whether these items are stronger related to the civic engagement or the social influence scale.

Page 14: In the text, it says that ingroup norms were measured with four items, but in the Appendix there are five.

Page 14: It would be helpful for replicability, if the R-packages that were used were listed. This is especially relevant as the formula to calculate McDonalds omega slightly varies across different R-packages.

Page 16: Table 3 is missing the *N*. In the first row of Table 3 "Item total correlation", etc. should not be in italics, only M and SD (same for Table 5; and please check the usage of italics across the whole manuscript). In addition, it is not clear whether the factor loadings presented here are standardized or not.

Page 18: In the last sentence of the first paragraph, it should say Table 3 and the syntax of the sentence is off.

Page 18: The last paragraph of Section 3.3 could be moved to Section 3.4 as it describes relations/correlations with other variables.

Page 20, line 556-557: The same applies to some items of the civic engagement scale (e.g., looking for allies or being in contact with a group also seem like preparatory behaviors).

Regarding open science, the authors preregistered the study, but I was not able to evaluate whether the description in the manuscript deviates from the preregistration as the link was blinded for review. Again, I think the manuscript has great potential and I commend the authors for the important sample and the work they have put into the manuscript so far (it flows nicely!). I hope that the number of comments does not discourage the authors and that they are helpful in improving the manuscript.
I am looking forward to reading a revised version of the manuscript,

Lukas Engel

Abrahamse, W., & Steg, L. (2013). Social influence approaches to encourage resource conservation: A meta-analysis. *Global environmental change*, *23*(6), 1773-1785. https://doi.org/10.1016/j.gloenvcha.2013.07.029

Bergquist, M., Thiel, M., Goldberg, M. H., & van der Linden, S. (2023). Field interventions for climate change mitigation behaviors: A second-order meta-analysis. *Proceedings of the National Academy of Sciences*, *120*(13), e2214851120. https://doi.org/10.1073/pnas.2214851120

Fritsche, I., Barth, M., Jugert, P., Masson, T., & Reese, G. (2018). A social identity model of pro-environmental action (SIMPEA). *Psychological review*, *125*(2), 245. https://doi.org/10.1037/rev0000090

Kaiser, F. G. (2021). Climate change mitigation within the Campbell paradigm: doing the right thing for a reason and against all odds. *Current Opinion in Behavioral Sciences*, *42*, 70-75. https://doi.org/10.1016/j.cobeha.2021.03.024

Kaiser, F. G., & Lange, F. (2021). Offsetting behavioral costs with personal attitude: Identifying the psychological essence of an environmental attitude measure. *Journal of Environmental Psychology*, *75*, 101619. https://doi.org/10.1016/j.jenvp.2021.101619

Kaiser, F. G., & Wilson, M. (2000). Assessing People's General Ecological Behavior: A Cross-Cultural Measure 1. *Journal of applied social psychology*, *30*(5), 952-978. https://doi.org/10.1111/j.1559-1816.2000.tb02505.x

Lalot, F., Räikkönen, J., & Ahvenharju, S. (2025). An Item Response Theory Approach to Measurement in Environmental Psychology: A Practical Example with Environmental Risk Perception. *Journal of Environmental Psychology*, 102520. https://doi.org/10.1016/j.jenvp.2025.102520

Lange, F. (2024). What is measured in pro-environmental behavior research?. *Journal of Environmental Psychology*, 102381. https://doi.org/10.1016/j.jenvp.2024.102381

Stern, P. C., Dietz, T., Abel, T., Guagnano, G. A., & Kalof, L. (1999). A value-belief-norm theory of support for social movements: The case of environmentalism. *Human ecology review*, 81-97. https://www.jstor.org/stable/24707060

Review B

The manuscript 'Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school student's civic engagement for climate protection' presents results from the psychometric testing of a measurement instrument for assessing students' civic engagement for climate protection (hereinafter referred to as "CECP"). Overall, the manuscript is very well comprehensible and the construct to be measured is highly relevant: the ability to contribute to sustainable development (and thus also to climate protection) together with others (which includes civic engagement) is at the center of most competence approaches to global education for sustainable development (OECD, 2018; UNESCO, 2017) - at the same time, there are not yet sufficient measures available for this area (Rieckmann, 2022). Education for sustainable development (ESD) and climate education often still focus too strongly on individual responsibility and neglect changing structures (Budde & Blasse, 2023). However, there are also substantial criticisms of the manuscript, the most fundamental of which concerns the process and presentation of instrument development. I will line out this aspect and some others,
followed by some minor issues, before I come to a conclusion.

Clarification of terms and references to models (Introduction/Background). In the presentation of CECP, I missed the reference to general civic engagement and civic competences, which already have a strong connection to sustainability issues (Abs et al., 2024). Furthermore, the relationship between CECP and the handprint is not clearly stated. In some places, the authors write that civic engagement "contributes" to the handprint (e.g., p.10, l. 277-279; p.22, l. 623-625). Elsewhere, the two terms are used synonymously (e.g., p.10, l. 272-274). In relation to the handprint, references to transformative learning (Germanwatch, 2023; Singer-Brodowski, 2016) and to instrumental vs. emancipatory ESD (e.g., Vare & Scott, 2007) would also be important.
The description of the Norm Activation Model is certainly sufficient for a readership familiar with environmental psychology. Otherwise, the information is too sparse to be able to understand the model. Furthermore, with regard to nomological validity, I missed the specifications of expectations with regard to the associations with related constructs.

Test Development Process (Methods). The test development process is barely described and therefore not transparent, nor are the underlying instruments presented. This makes it difficult for the reader to understand where the actual development took place - was it mainly in the selection of items suitable for students, was it in the addition of examples? What adaptations or additions have been made? It was also not described whether there were any pilot

tests, for example in the form of think-aloud protocols, to test understanding of the items. For example, I am surprised that item 7 (p.12), unlike some others, is not explained with an example - it could be difficult to assign concrete behaviours to 'I tried to make climate-friendly behaviour easier or more approachable in my environment' without further explanation. Apart from that, item 8 ["I initiated changes in the overall conditions of my environment (e.g., vegan options in the canteen) towards more climate protection"] and the associated example could also be a way to make climate-friendly behaviour (□ Item 7) easier – Possibly, the obtained free text answers that can be matched to the existing items can not only be interpreted as evidence of validity, on the contrary, they could even be an indication that the students did not realize which activities were specifically meant when answering the Likert-type items. A central point of criticism is the following: the response categories (6-point Likert scale ranging from 1 = not true at all to 6 = completely true) do not match the items: The question on specific actions actually carried out can only be answered with yes and no. If a higher variability in the answers is desired, a frequency scale could have been used. As these aspects of the development process cannot be adjusted in the present study, the associated limitations should be emphasized in the discussion.

Analysis (Methods). The authors' argument regarding the general use of the instrument "as a measure of individual differences between adolescents irrespective of their grouping in clusters" (P.14 l.396-398) as a justification for not taking into account the multi-level structure of the data does not seem convincing to me. Especially for a construct such as the CECP, the investigation of effects at different levels, e.g. peer group effects, seems to be of interest. D'Haenens and colleagues (2010), for example, warn of biases if the multi-level structure is ignored when using EFA. The authors should reconsider using multilevel-EFA (see Muthén & Muthén, 1998-2017).

Discussion. The authors acknowledge that the present study is to be regarded as a pilot test and that further studies are needed to finalize the instrument. This should be emphasized more strongly, for example by discussing the weaknesses of the development process and instrument characteristics and by using an even more cautious language with regard to the results - for example, rather than referring to evidence of one-dimensionality (see p.22, l. 631-632), the results should be interpreted as initial indications on the construct's structure.

Open Science. As far as aspects of Open Science are concerned, this study can be considered largely exemplary, as the study was pre-registered and the analysis code is publicly accessible.
However, I do not understand the argument that the data collected is too sensitive for publication. All information that could allow conclusions on individuals, such as school affiliation and possibly individual socio-demographic data, could be removed in the public file. However, publication for secondary use is only possible if special consent has been obtained for this in the declarations of consent (Verbund Forschungsdaten Bildung, 2019).

Minor issues:
p.1 lines 1-3: 'Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school student's civic engagement for climate protection' students' instead of student's
p. 6 line 205: '1.4' is repeated
p. 10 lines 286 - 87 'Ratings of three and lower indicate a somewhat negative response and ratings of four and higher indicate a somewhat positive response' - I think this explanation is not necessary and can be omitted
p.11 lines 300-305 The explanation of how 'my environment' is explained would fit better in a footnote

p. 21 lines 586 - 87 the individual 'psychosocial variables' should be specifically mentioned again here
In summary, the instrument presented in the manuscript still needs to be further developed and validated. However, interested researchers, not only from the field of environmental psychology, but also from empirical educational research on ESD, should have access to the instrument. I therefore support the publication of this paper, provided that the aforementioned revisions (especially a detailed description of the test development process and a discussion of limitations) are implemented.

References

Abs, H. J., Hahn-Laudenberg, K., Deimel, D., & Ziemes, J. F. (Eds.). (2024). *ICCS 2022: Schulische Sozialisation und politische Bildung von 14-Jährigen im internationalen Vergleich*. Waxmann.

Budde, J., & Blasse, N. (2023). Bildung für nachhaltige Entwicklung zwischen Programmatik und Praxis. *ZEP – Zeitschrift Für Internationale Bildungsforschung Und Entwicklungspädagogik*, *2023*(2), 4–9. https://doi.org/10.31244/zep.2023.02.02

D'Haenens, E., Van Damme, J., & Onghena, P. (2010). Multilevel exploratory factor analysis: Illustrating its surplus value in educational effectiveness research. *School Effectiveness and School Improvement*, *21*(2), 209–235. https://doi.org/10.1080/09243450903581218

Germanwatch e.V. (2023). *Methodenhandbuch Handabdruck: Eine Orientierungshilfe für transformative Bildung und Engagement*. Germanwatch e.V.https://www.germanwatch.org/sites/default/files/germanwatch_methodenhandbuch-handabdruck_2023.pdf

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (Eighth Edition). Muthén & Muthén. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

OECD. (2018). *Preparing our youth for an inclusive and sustainable world: The OECD PISA global competence framework*. OECD Publishing. http://www.oecd.org/pisa/aboutpisa/Global-competency-for-an-inclusive-world.pdf

Rieckmann, M. (2022). Developing and Assessing Sustainability Competences in the Context of Education for Sustainable Development. In G. Karaarslan-Semiz (Ed.), *Education for Sustainable Development in Primary and Secondary Schools* (pp. 191–203). Springer. https://doi.org/10.1007/978-3-031-09112-4_14

Singer-Brodowski, M. (2016). Transformatives Lernen als neue Theorie-Perspektive in der BNE. In *Umweltdachverband GmbH (Hrsg.): Jahrbuch Bildung für nachhaltige Entwicklung – Im Wandel. Forum Umweltbildung im Umweltdachverband* (pp. 130–139).

UNESCO. (2017). *Education for Sustainable Development Goals: Learning Objectives*. UNESCO. https://www.unesco.de/sites/default/files/2018-08/unesco_education_for_sustainable_development_goals.pdf

Vare, P., & Scott, W. (2007). Learning for a change: Exploring the relationship between education and sustainable development. *Journal of Education for Sustainable Development*, *1*(2), 191–198.

Verbund Forschungsdaten Bildung. (2019). *Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen* (p. 7 pages). DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation : Frankfurt am Main. https://doi.org/10.25656/01:22297

Environmental Psychology Open

Authors' Response to the Reviewers Round 1

Authors' Response to the Editor

Dear Dr. Mathias Hofmann,

thank you for the opportunity to submit a revised version of our manuscript, titled "Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school students' civic engagement for climate protection". We sincerely appreciate the time and effort you and the reviewers have invested in providing valuable feedback.

We have carefully considered the reviewers' suggestions and have incorporated revisions to address their recommendations. To facilitate the review process, all changes in the revised manuscript are highlighted in red font. Below, we provide a point-by-point response to the reviewers' comments and questions.

Thank you for your consideration.

Authors' Response to comments from Reviewer A

First, I thank the authors and the editors for giving me the opportunity to read this interesting manuscript. The manuscript presents a validation study for a newly developed scale to measure civic engagement for climate protection among adolescents. In a sample of $N = 360$ secondary school students, the authors find evidence for the unidimensionality of the scale, its high reliability, and the expected relations with theoretically derived psychological constructs and other classes of behavior. I think that this manuscript can make an important contribution based on its focus on an understudied population and an often-overlooked behavioral type in environmental psychology. However, I believe that this manuscript should be improved before publication. Below, I have listed some major and minor concerns.

Response: We appreciate your detailed and positive feedback on our manuscript. Below, we provide a point-by-point response to your comments, outlining the revisions we have made to the manuscript.

Concern 1

The abstract gives a nice overview of the study and its findings. But the last statement is too general and should mention the limitation that the scale has only been tested with German students/adolescents so far. Currently, the abstract suggests general applicability. As a minor
note, M_{age} is formatted differently in the German and English abstract.

Response: Thank you for pointing this out. We have revised the final statement of the abstract and added a statement that the scale should be validated in school contexts of other countries. We have also ensured that M_{age} is formatted consistently across both the German and English abstracts.

Environmental Psychology Open

p.2, l. 51: Results indicate applicability of the scale to measure civic engagement for climate protection, warranting replication in school contexts of other countries.

Concern 2

Turning to the main text, I see two major issues: one is theoretical and the other methodological. Starting with the theoretical issue, civic engagement for climate protection is not clearly enough differentiated from social influence and collective action (some aspects of this are mentioned in the discussion, but it would benefit the reader, if this was moved to Section 1.1). The authors nicely derive how civic engagement is different from individual consumption, but it would be helpful to clearly define the relation between civic engagement and collective action and social influence (in the paragraph that introduces civic engagement). Throughout the manuscript, I struggled with distinguishing between the different types, for example, as "collective efforts" is part of the current definition of civic engagement, and Item 7 of the civic engagement scale and Item 4 of the social influence scale overlap. In addition the authors speak of wanting to understand how people advocate for climate protection in the first sentence of Section 1.2 which would also fit the description of social influence (e.g., Abrahamse & Steg, 2013).

Response: Thank you for your comment. We clarified the differentiation between social influence and civic engagement in section 1.1 and added that we understand civic engagement and collective action synonymously. We changed the ambiguous wording in the first sentence of section 1.2.

p.4, l. 124: Another form of climate protection behavior is to influence close others to change their behavior within their scope of action and in turn reduce their ecological footprint, for example by convincing a family member to cycle to work instead of driving by car (Hamann & Masson, 2020).

p.4, l. 134: A third form of climate protection behavior is civic engagement for climate protection, also referred to as collective action (Hamann & Masson, 2020; Kranz et al., 2022).

p.5, l. 178: To understand how people show civic engagement for climate protection, it is essential to consider a variety of behaviors (Tindall et al., 2003).

Concern 3

Besides this definitional aspect, I would question what construct the authors measure with the civic engagement scale. The reasoning in the introduction is coherent, however in Section 1.1 it is unclear whether the scale is supposed to measure a person's ecological handprint (person property; see Lange, 2024) or a property of the behavior (e.g., frequency of specific behaviors). I highly recommend reading Lange (2024) to determine what exactly they are measuring and to adjust the definition accordingly. If the authors come to the conclusion that they measure a person property, I would question whether this property (i.e., the intention to protect the climate by "focusing on systemic causes of environmental problems and the promotion of environmental sustainability through collective efforts") is different from the intention to protect the climate by reducing personal consumption or influencing others. Previous work shows that these different behavioral types might have the same underlying intention (e.g., Kaiser, 2021; Kaiser & Lange, 2021). The high correlations between the different behavior self-report scales (even in light of restricted variability and reliability) hint towards this interpretation. I think the point you are making in the first paragraph on Page 6 is really interesting: that different forms of civic engagement might just represent different levels of a unidimensional construct. This could very well also be the case for different forms of climate-protective behaviors.

Environmental Psychology Open

Response: Thank you for suggesting reading Lange (2024). We apologize for the imprecision in our description of the relationship between civic engagement for climate protection and the socio-ecological handprint. The socio-ecological handprint is an indicator of impact, to which civic engagement for climate protection might contribute. We have changed the wording in the manuscript to consistently refer to civic engagement contributing to the socio-ecological handprint. With our scale, we aim to measure self-reported agreement to the performance of several exemplary behaviors reflecting a person's civic engagement for climate protection, thus, behaviors reflecting a common factor. However, we argue our scale does not measure intention as the items are worded retrospectively and not prospectively. The question of whether civic engagement for climate protection scale means are stable within persons over time lies beyond the scope of this cross-sectional research, but should definitely be addressed in future research. Both the theoretical model (Hamann &Masson, 2020) that guided our scale development as well as the empirical results we obtained indicate that civic engagement for climate protection, social influence, and individual consumption can be distinguished (see also Concern 7).

Concern 4

When it comes to the methodology, I do not think that Classical Test Theory (CTT) is the most appropriate methodological approach to answer the research questions (especially, as the authors explicitly want to evaluate item characteristics), and I would suggest several additional analyses before publication of this manuscript. It is not clear to me why the authors used CTT to evaluate item characteristics when Item Response Theory (IRT) would have clear advantages in this regard (e.g., Lalot, 2025). I would strongly suggest to test the construct dimensionality and evaluate the scale using IRT methods like the Partial-Credit model, Graded Response model or dichotomous Rasch model (would necessitate prior item dichotomization). If you stick to CTT, I would expect an explanation why it was chosen over IRT.

Response: Thank you for pointing out the advantages of Item Response Theory. We conducted further analyses of item characteristics and dimensionality by applying a Rasch rating scale model and reported these results, please see pages 15 and 16.

Concern 5

Furthermore, the authors are unclear about the evaluation of the correlations for Research Question 4: At what *r*-level are the correlations in line with the assumptions (e.g., is a strong correlation not in line with discriminant validity and a weak correlation not in line with the appropriateness of the nomological network)? In addition, it might be interesting to investigate how the different behavior classes vary in their relations with the psychosocial predictors (if they are not found to represent the same psychological property).

Response: Thank you for your comment. In research question 4, we state that theoretically proposed bivariate positive correlations between civic engagement for climate protection and perceived knowledge about options for civic engagement for climate protection, perceived risk of climate change, ingroup identification, ingroup norms, self- and collective efficacy beliefs are investigated as part of the nomological validity. We acknowledge that we did not specify the expected magnitude of association, however, we did not have further assumptions regarding the magnitude of the correlations. Further, based on the underlying theoretical models we did not specifically assume different relations between psychosocial predictors and forms of climate protection behaviors. Thus, a detailed analysis of relations goes beyond the scope of this work.

Environmental Psychology Open

Concern 6
For a thorough scale validation and evaluation, I would also expect the following tests: Does the scale measure the construct equally across groups (measurement invariance or differential item functioning); for example, I would expect some items to be differently understood by 17-year-olds vs. 12-year-olds, which could compromise the scale fairness?

Response: Thank you for raising this point, we agree that differential item functioning might be an issue. We tested measurement invariance across age groups and reported the results in 3.4.

p. 18, l. 544 : To assess whether the scale measures civic engagement for climate protection equally across age groups, we estimated a multiple-group Rasch model with five age groups (12 years, 13 years, 14 years, 15 years and ≥16 years), combining school students 16 years or older into one group due to the small number of older students. Root mean square differences (RMSD) quantify the distance between group-specific item characteristic curves and joint item characteristic curves. Differential item functioning (DIF) must be assumed if RMSD exceeds .12 (Baghaei & Robitzsch, 2025). None of the items were flagged for differential item functioning in either age group. Thus, measurement invariance across age can be assumed.

Concern 7
Going beyond HTMT, it should be tested whether the three behavior self-report scales (individual consumption, social influence, civic engagement) can be represented in a one-dimensional measurement models (either with EFA, CFA, or preferably a IRT method).

Response: Thank you for your suggestion. We assessed whether the three scales measuring climate protection behaviors (civic engagement, social influence, individual consumption) can be represented in a one-dimensional measurement model using principal component analysis on Rasch residuals and reported the results in 3.4.

p. 19, l. 563: Principal component analysis on Rasch residuals was performed on a Rasch rating scale model of all items included in the civic engagement, individual consumption and social influence scales. The first component of the principal component analysis on Rasch residuals had an eigenvalue of 2.11, which exceeds the upper threshold of 2.0 suggested to meet the assumption of one-dimensionality (Linacre, 2022).

Concern 8
The authors should evaluate the dimensionality and measurement models of all scales they use as the relation between the civic engagement scale and the other scales is such an integral part of this manuscript. Also, in case the authors stick to CTT, they should compute CFAs, which would allow to evaluate the model fit based on indices like CFI, RMSEA, and SRMR

Response: Thank you for your suggestion. We present results on the dimensionality and measurement models of all scales in the supplementary material (Table B9).

Environmental Psychology Open

Concern 9

In addition to these two issues, I have the following remarks: In Section 1.3, it is not clear why the authors chose these two specific theories for the nomological network. This should be made more explicit. Currently, they only state that both individual and collective variables seem to be important. But why did they then specifically choose the NAM and not the Value-Belief-Norm theory (Stern et al., 1999), for example, for the individual factors? I am not saying that the choice of theories is bad, just that it should be made more explicit why they were chosen for this study. Also, the SIMPEA is presented as if the focus of it is on civic engagement for climate action, when this is not the case (Fritsche et al., 2018).

Response: Thank you for giving us the opportunity to clarify this. We have made the reasoning for the choice of theories more explicit and adjusted the wording to clarify that SIMPEA focuses on climate protection behavior as a whole.

p. 6, l. 215: Individual-level predictors of civic engagement for climate protection can be derived from the Norm Activation Model (Schwartz & Howard, 1981). The Norm Activation Model was developed to predict prosocial behavior and is commonly applied in research on climate protection behavior (Klöckner, 2013). The Norm Activation Model is relevant for civic engagement for climate protection because it accounts for the uptake of behaviors that express a sense of responsibility for others, even in the face of personal cost (Seger & Böcker, 2023).

p. 6, l. 231: The Social Identity Model of Pro-Environmental Action (SIMPEA, Fritsche et al., 2018) is based on the Social Identity Model of Collective Action (SIMCA, van Zomeren et al., 2008). SIMPEA focuses on collective variables that are associated with climate protection behavior. It has been successfully applied to explain civic engagement for climate protection amongst young people (Wallis & Loy, 2021). SIMPEA proposes that climate protection behavior is determined by ingroup identification, ingroup norms, and collective efficacy beliefs.

Concern 10

There are no explicit hypotheses in this manuscript which makes the evaluation of the research questions difficult. Hypotheses could be added to improve understandability, for example in Section 1.5.

Response: Thank you for pointing this out. We agree that hypotheses might improve comprehension, however, we did not specify explicit hypotheses due to the exploratory nature of the analysis of a new measure.

Concern 11

In general, I was wondering whether difficult terms like petition were explained? Was the understanding of items and introductory text checked? To me the wording seems quite complex for the younger participants of the study (e.g., Verankerung, Verbündete, Petition). Were there any pretests to ensure understandability? In addition, asking a 12-year-old whether they donated money might not be suitable. These aspects of the study should be discussed as they might have affected the findings.

Response: Thank you for raising these concerns. We did not conduct pilot tests with adolescents to ensure comprehension of the items. During the scale development, we did discuss if donating is relevant for school students in our sample. However, donating was found to be a somewhat relevant behavior in 14-22 year-olds in a representative survey by the German Ministry for the Environment (Bundesministerium für Umwelt, 2022)

and was thus included in the scale. We added remarks on this in section 2.4 and the discussion.

p. 10, l. 325: Comprehensibility of the scale for the target group was checked by two school teachers, however, no pilot data from school students was collected.

p. 21, l. 630: Item 5 had the lowest correlations with the other scale items. In contrast to the other civic engagement behaviors, donating is based on monetary (vs. time) resources which limits accessibility for adolescents (Han et al., 2023).

p. 23, l. 725: A further limitation is the lack of pilot testing with members of the target group.

Concern 12
Regarding the results, the part-whole correlations were not described in Section 2.4 and there is neither a mention nor a reference of the acceptable range.

Response: Thank you for addressing this. We added this in 2.4 and cited the acceptable range in 3.1.

p. 14, l.424: To address research question 1, item characteristics were evaluated using elements of classical test theory (means, standard deviations, item total correlations, skewness and kurtosis) and item response theory (Rasch measurement).

p. 15, l.463: Part-whole corrected item-total correlations were positive and within the recommended range between .40 - .70 (Moosbrugger & Kelava, 2007) for all items (between .49; item 3 and .69; item 8), which indicates that each item makes an appropriate contribution to the construct to be measured by the instrument as a whole.

Concern 13
Furthermore, there is a nice description of how missing values were treated on the civic engagement scale, but there is no mention of how missing data on all other constructs were treated and distributed. This would be important information to present.

Response: Thank you for pointing this out. We expanded this section to cover missing data in all relevant variables.

p. 15, l. 456: Missing data was rare for all scales (range: 0% - 1.93% missing values per item) and tested to be missing completely at random ($\chi 2\,(948) = 990$, $p = .17$) (Little, 1988). Missing values were imputed using predictive mean matching (Goretzko, 2021).

Concern 14
I understand the reasoning by the authors that civic engagement behavior is rarely shown and that either a dichotomous scale or frequency scale would have led to restricted variability. But at the same time creating a 6-point Likert-scale for a truly dichotomous variable (behavior shown yes-no) can also affect the reliability of the responses (Kaiser & Wilson, 2000). This could be discussed or mentioned.

Environmental Psychology Open

Response: Thank you for your comment. In the development of the scale, we deliberately formulated the items in such a way that they do not require dichotomous answers, but that gradual agreement is possible (e.g., by always using the plural, i.e. contacts, petitions, etc). We do acknowledge that other response scales might be appropriate and addressed this point in the discussion.

p. 20, l. 603: The 6-point response scale may have not matched the actual occurrence of and variability in behavior. Depending on the scope of research, future studies could consider to alternatively choose a binary response scale (i.e., behavior shown vs. not shown) or a frequency scale (e.g., behavior frequency per month).

Concern 15
When it comes to the sample description it would improve readability to either define social status already in the corresponding section (i.e., what was the item) or refer to the section where it is described. Generally, it would be an important information whether the assessed scales for the variables in the nomological network were previously validated in a similar sample. For example, I would assume that the social ladder scale might be perceived a lot differently by children/adolescents compared to adults (e.g., more volatile).

Response: Thank you for your comment. We added a reference to the section where the assessment of social status is described in the sample description. Further, we added the limitation that many scales used to describe the nomological network were developed for the overarching study project and have not yet been validated for the sample of school students.

p. 8, l. 304: The mean perceived social status (see 2.3) of participants was 4.11 (SD = 1.34) on a scale of 1-10.

p. 23, l. 721: It should be noted that many of the scales used to describe the nomological network were developed for the Public Climate School study project. They have not yet been validated in the relevant sample and their relation to civic engagement for climate protection has not been tested previously.

Concern 16
On page 20, it is written that "Item 7 ("tried to make climate-friendly behavior easier or more approachable in my environment") received significantly higher ratings than the other items." I am wondering whether this significance was tested. If yes, the corresponding test statistic should be presented.

Response: Thank you for raising this concern. We added the test statistic in section 3.1.

p. 15, l. 479: Ratings for item 7 were significantly higher than ratings for the other items ($W = 28727$, $p < .001$).

Concern 17
The discussion of item 5 is missing that donating is probably the least accessible behavior for children and adolescents which could have the observed negative effect on the correlations.

Environmental Psychology Open

Response: Thank you for this suggestion. We added a remark on this.

p. 21, l. 632: Item 5 had the lowest correlations with the other scale items. In contrast to the other civic engagement behaviors, donating is based on monetary (vs. time) resources which limits accessibility for adolescents (Han et al., 2023).

Concern 18
The discussion of the nomological network should be expanded. Identifying knowledge gaps as an effective point to intervene is a bit counterintuitive given a) the small effects of information interventions (e.g., Bergquist et al., 2023), and b) that subjective knowledge in your study has (descriptively) the weakest correlation with civic engagement.

Response: Thank you for pointing this out. We have revised the discussion of the nomological network.

p. 22, l. 683: The positive relationships between the civic engagement measure and perceived knowledge about options for civic engagement for climate protection, perceived risk of climate change, ingroup identification, ingroup norms, self- and collective efficacy beliefs were in accordance with theoretical expectations from the Norm Activation Model and the SIMPEA model and show starting points through which civic engagement for climate protection may be addressed in interventions. Positive relations between the civic engagement scale and collective efficacy beliefs, ingroup norms, and especially ingroup identification underline the significance of social aspects and peer interactions for the civic engagement of adolescents (Fritsche et al., 2018; Wallis & Loy, 2021). Further, the perceived risk of climate change, perceived knowledge about options for civic engagement for climate protection and self-efficacy beliefs may be strengthened through action-oriented and empowering education for sustainable development. These findings emphasize the role of schools as places for education, formation of social bonds, and opportunities to show civic engagement for climate protection for school students.

Concern 19
Finally, I am missing in the discussion a) a CAVEAT for the low reliability of the individual consumption scale (which could be taken into account in the correlation with civic engagement by computing error-attenuation corrected correlations), and b) the possibility that the negative correlation between age and civic engagement might be rooted in younger students not having understood some of the items.

Response: Thank you for these suggestions. We added these points in the discussion.

p. 21, l. 666: The moderate positive association between civic engagement and individual consumption behaviors should be interpreted cautiously due to the low reliability of the individual consumption scale, but is in line with previous studies (Alisat & Riemer, 2015; Reuter & Frick, 2024).

p. 21, l. 660: A possible explanation might be that younger students in our sample were more likely to display higher social desirability (Ng et al., 2024) or did not comprehend the items fully.

Environmental Psychology Open

Minor comments:

There is no in-text reference to Figure 1.

Response: We added an in-text reference (p. 5, l. 160)

Page 6: 1.4 is two times in the heading

Response: We corrected this. (p. 7, l. 249)

Page 8 and 9: The categories for municipality size in the text do not match the categories in the table (100,000-1,000,000 vs 100,000-999,999); in the table the groups are missing a data point: a size of 1,000,000 is not represented.

Response: We corrected this in the table (p. 9)

Page 9: The statistical abbreviations M, SD and n should be in italics.

Response: We corrected this (p. 9).

Page 10 (table note): Please explain why there are missing values / why the numbers not always add up to the total sample size. Are the percentages based on the total sample size or the *sample* size that provided an answer to the item?

Response: Apologies for this oversight. We added the number of missings for the relevant variables and corrected percentages to refer to the percentage of the sample size. (p. 9)

Page 13: The item for knowledge measures subjective knowledge and should be labelled as such. Currently, one might get the impression that a knowledge test was assessed when not reading the measures section.

Response: We named it "perceived knowledge" throughout the manuscript.

Page 13: This cannot changed, but the scale for measuring self-efficacy beliefs rather measures beliefs about the ability for social influence and not civic engagement in my interpretation of the items ("Umfeld" as social environment). This just highlights how important a clear definition and differentiation from related constructs is which should be added in the theoretical background. It would be interesting to compute whether these items are stronger related to the civic engagement or the social influence scale.

Environmental Psychology Open

Response: Thank you for this observation. With "Umfeld" in these items we referred back to the Introduction text of the civic engagement scale, in which the school, clubs, religious communities and place of residence are mentioned. However, we do agree that the wording is ambiguous and not specific to civic engagement for climate protection. We addressed this in the limitations.

p. 23, l. 721: It should be noted that many of the scales used to describe the nomological network were developed for the Public Climate School study project. They have not yet been validated in the relevant sample and their relation to civic engagement for climate protection has not been tested previously.

Page 14: In the text, it says that ingroup norms were measured with four items, but in the Appendix there are five.

Response: Apologies for this oversight. We had mistakenly included an item in the table in the Appendix that was not part of this scale and have now corrected this (supplemental material B7). This mistake changes nothing regarding our data analyses.

Page 14: It would be helpful for replicability, if the R-packages that were used were listed. This is especially relevant as the formula to calculate McDonalds omega slightly varies across different R-packages.

Response: Thank you for this important suggestion. We added the main R-Packages used to section 2.4.

p. 14, l. 423: Data analysis was performed using R (RCoreTeam, 2024) and the packages psych (Revelle, 2024) and TAM (Robitzsch et al., 2024).

Page 16: Table 3 is missing the *N*. In the first row of Table 3 "Item total correlation", etc. should not be in italics, only M and SD (same for Table 5; and please check the usage of italics across the whole manuscript). In addition, it is not clear whether the factor loadings presented here are standardized or not.

Response: We clarified that we reported standardized factor loadings, added the sample size in the notes and corrected the usage of italics (p. 16 and p. 19).

Page 18: In the last sentence of the first paragraph, it should say Table 3 and the syntax of the sentence is off.

Response: We corrected this.

Page 18: The last paragraph of Section 3.3 could be moved to Section 3.4 as it describes relations/correlations with other variables.

Environmental Psychology Open

Response: We have moved the paragraph (p. 18. l. 535)

Page 20, line 556-557: The same applies to some items of the civic engagement scale (e.g., looking for allies or being in contact with a group also seem like preparatory behaviors).

Response: We revised the phrasing to reflect our reasoning more precisely.

p. 21, l. 649: However, educating oneself alone does not contribute to a person's socio-ecological handprint without subsequent engagement. We consider it as an expression of contemplation rather than goal-oriented preparation or enactment of civic engagement for climate protection.

Regarding open science, the authors preregistered the study, but I was not able to evaluate whether the description in the manuscript deviates from the preregistration as the link was blinded for review. Again, I think the manuscript has great potential and I commend the authors for the important sample and the work they have put into the manuscript so far (it flows nicely!). I hope that the number of comments does not discourage the authors and that they are helpful in improving the manuscript.

I am looking forward to reading a revised version of the manuscript, Lukas Engel

Response: Thank you for taking the time to review our manuscript!

Authors' Response to comments from Reviewer B

Review of the manuscript „Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school student's civic engagement for climate protection" The manuscript 'Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school student's civic engagement for climate protection' presents results from the psychometric testing of a measurement instrument for assessing students' civic engagement for climate protection (hereinafter referred to as "CECP"). Overall, the manuscript is very well comprehensible and the construct to be measured is highly relevant: the ability to contribute to sustainable development (and thus also to climate protection) together with others (which includes civic engagement) is at the center of most competence approaches to global education for sustainable development (OECD, 2018; UNESCO, 2017) - at the same time, there are not yet sufficient measures available for this area (Rieckmann, 2022). Education for sustainable development (ESD) and climate education often still focus too strongly on individual responsibility and neglect changing structures (Budde & Blasse, 2023). However, there are also substantial criticisms of the manuscript, the most fundamental of which concerns the process and presentation of instrument development. I will line out this aspect and some others, followed by some minor issues, before I come to a conclusion.

Response: We appreciate the constructive feedback on our manuscript. Below, we provide a detailed response to each point and describe the revisions we have made accordingly.

Environmental Psychology Open

Clarification of terms and references to models (Introduction/Background).

Concern 1
In the presentation of CECP, I missed the reference to general civic engagement and civic competences, which already have a strong connection to sustainability issues (Abs et al., 2024).

Response: Thank you for this observation, we included an integration of civic engagement for climate protection within the context of general civic engagement.

p. 4, l. 163: Civic engagement describes an individual's behaviors that express a sense of responsibility for and identification with humanity in a global context, one component of which is climate protection (Abs et al., 2024).

Concern 2
Furthermore, the relationship between CECP and the handprint is not clearly stated. In some places, the authors write that civic engagement "contributes" to the handprint (e.g., p.10, l. 277-279; p.22, l. 623 625). Elsewhere, the two terms are used synonymously (e.g., p.10, l. 272-274).

Response: Apologies for this imprecision. We changed the wording to consistently refer to civic engagement contributing to the socio-ecological handprint (p. 10, l. 316; p. 21, l. 650; p. 23, l. 729)

Concern 3
In relation to the handprint, references to transformative learning (Germanwatch, 2023; Singer-Brodowski, 2016) and to instrumental vs. emancipatory ESD (e.g., Vare & Scott, 2007) would also be important.

Response: Thank you for this comment. We have added references to transformative learning and education for sustainable development.

p. 4, l. 153: The NGO Germanwatch has further developed the concept and established the focus on structural change and transformative learning (Heitfeld & Reif, 2020; Reif & Heitfeld, 2015).

p.4, l. 156: A comprehensive education for sustainable development imparts competencies on both the implementation of proven effective climate protection behaviors in one's own life and critical reflection on which changes are relevant for climate protection at the societal level (Vare & Scott, 2007).

Concern 4
The description of the Norm Activation Model is certainly sufficient for a readership familiar with environmental psychology. Otherwise, the information is too sparse to be able to understand the model.

Response: Thank you for raising this concern, we agree that in our initial submission we did not sufficiently describe the Norm Activation Model. We have elaborated our description of the Norm Activation Model.

p.6, l. 215: Individual-level predictors of civic engagement for climate protection can be derived from the Norm

Environmental Psychology Open

Activation Model (Schwartz & Howard, 1981). The Norm Activation Model was developed to predict prosocial behavior and is commonly applied in research on climate protection behavior (Klöckner, 2013). The Norm Activation Model is relevant for civic engagement for climate protection because it accounts for the uptake of behaviors that express a sense of responsibility for others, even in the face of personal cost (Seger & Böcker, 2023). Based on the Norm Activation Model, civic engagement for climate protection is driven by personal norm: a sense of personal responsibility to act. Multiple prerequisites must be met, so that an individual's personal norm to enact civic engagement for climate protection is activated (Klöckner, 2013). First, the individual must be aware of a social problem, in this case the risk that climate change poses on humanity. Further, an individual must be aware of the consequences of their own behavior in the context of this problem, e.g., knowledge on which behaviors are effective for climate protection. Lastly, an individual must assume responsibility and feel competent to perform these behaviors (Klöckner, 2013).

Concern 5

Furthermore, with regard to nomological validity, I missed the specifications of expectations with regard to the associations with related constructs.

Response: Thank you for pointing this out. In research question 4, we stated that bivariate positive correlations between civic engagement for climate protection and perceived knowledge about options for civic engagement for climate protection, perceived risk of climate change, ingroup identification, ingroup norms, self- and collective efficacy beliefs are in line with the theory and therefore an indication of nomological validity. We acknowledge that we did not specify the expected magnitude of association, however, we did not have further assumptions regarding the magnitude of the correlations.

Test Development Process (Methods).

Concern 6

The test development process is barely described and therefore not transparent, nor are the underlying instruments presented. This makes it difficult for the reader to understand where the actual development took place - was it mainly in the selection of items suitable for students, was it in the addition of examples? What adaptations or additions have been made? It was also not described whether there were any pilot tests, for example in the form of think-aloud protocols, to test understanding of the items. For example, I am surprised that item 7 (p.12), unlike some others, is not explained with an example - it could be difficult to assign concrete behaviours to 'I tried to make climate-friendly behaviour easier or more approachable in my environment' without further explanation.

Reponse: Thank you for pointing this out. We elaborated on the development process in section 2.2.

p. 10, l. 312: A scale for the assessment of civic engagement for climate protection in secondary school students was developed. Two of the items (item 2 and item 6) were adopted from the 'communication and engagement behavior' scale developed for the Public Climate School study 2021 (Keller et al., 2024). The other items were newly developed and reflected behaviors that contribute to the socio-ecological handprint based on work by Reif and Heitfeld (2015), Heitfeld & Reif (2020), and Hamann and Masson (2020) by members of blinded for review. In an iterative process comprising multiple online meetings, item selection and phrasing were refined with

Environmental Psychology Open

the feedback and suggestions of researchers and practicians in education for sustainable development (blinded for review). Items were selected with the aim of representing behaviors that are attainable and comprehensible for the target group of secondary school students. To this end, examples from the school context for two items (item 3 and item 7) and an extensive introductory text (see supplementary material A) prefacing the scale were added. Comprehensibility of the scale for the target group was checked by two school teachers, however, no pilot data from school students was collected.

Concern 7

Apart from that, item 8 ["I initiated changes in the overall conditions of my environment (e.g., vegan options in the canteen) towards more climate protection"] and the associated example could also be a way to make climate-friendly behaviour (→ Item 7) easier – Possibly, the obtained free text answers that can be matched to the existing items can not only be interpreted as evidence of validity, on the contrary, they could even be an indication that the students did not realize which activities were specifically meant when answering the Likert-type items.

Response: Thank you for pointing this out. We added this point to the discussion.

p. 21, l. 646: On the other hand, those responses might suggest that school students had difficulties in assigning their activities to the items.

Concern 8

A central point of criticism is the following: the response categories (6-point Likert scale ranging from 1 = not true at all to 6 = completely true) do not match the items: The question on specific actions actually carried out can only be answered with yes and no. If a higher variability in the answers is desired, a frequency scale could have been used. As these aspects of the development process cannot be adjusted in the present study, the associated limitations should be emphasized in the discussion.

Response: Thank you for your comment. In the development of the scale, we deliberately formulated the items in such a way that they do not require dichotomous answers, but that gradual agreement is possible (e.g., by always using the plural, i.e. contacts, petitions, etc). We do acknowledge that other response scales might be an alternative and addressed this point in the discussion.

p. 20, l. 603: The 6-point response scale may have not matched the actual occurrence of and variability in behavior. Depending on the scope of research, future studies could consider to alternatively choose a binary response scale (i.e., behavior shown vs. not shown) or a frequency scale (e.g., behavior frequency per month).

Concern 9

Analysis (Methods). The authors' argument regarding the general use of the instrument "as a measure of individual differences between adolescents irrespective of their grouping in clusters" (P.14 l.396-398) as a justification for not taking into account the multi-level structure of the data does not seem convincing to me. Especially for a construct such as the CECP, the investigation of effects at different levels, e.g. peer group effects, seems to be of interest. D'Haenens and colleagues (2010), for example, warn of biases if the multi-level structure is ignored when using EFA. The authors should reconsider using multilevel-EFA (see Muthén & Muthén, 1998-2017).

Environmental Psychology Open

Response: Thank you very much for raising this important point. As recommended in D'Haenens et al. (2010), we calculated intraclass correlation coefficients (ICCS) as a measure of the proportion of total variance that can be attributed to between-class or between-school group differences. In a three-level model, the ICC for the school level was 0 and the ICC for the class level was 0.087, indicating that 91% of variance is explained at the individual (i.e., student) level. D'Haenens et al. (2010) did only consider items with a "sufficient amount of variation between schools" (p. 216) in their multilevel-EFA. Items with ICCS smaller than 0.8 were not considered (D'Haenens et al., 2010). Based on this, we decided to conduct EFA on the individual level, in which the majority of variance occurred.

Concern 10

Discussion. The authors acknowledge that the present study is to be regarded as a pilot test and that further studies are needed to finalize the instrument. This should be emphasized more strongly, for example by discussing the weaknesses of the development process and instrument characteristics and by using an even more cautious language with regard to the results - for example, rather than referring to evidence of one-dimensionality (see p.22, l. 631-632), the results should be interpreted as initial indications on the construct's structure.

Reponse: Thank you for this suggestion, we agree that our language did not reflect the exploratory nature of our finding sufficiently in our initial submission. We have revised the discussion to reflect this more clearly.

Concern 11

Open Science. As far as aspects of Open Science are concerned, this study can be considered largely exemplary, as the study was pre-registered and the analysis code is publicly accessible. However, I do not understand the argument that the data collected is too sensitive for publication. All information that could allow conclusions on individuals, such as school affiliation and possibly individual socio-demographic data, could be removed in the public file. However, publication for secondary use is only possible if special consent has been obtained for this in the declarations of consent (Verbund Forschungsdaten Bildung, 2019).

Response: We appreciate the perspective on data sharing and agree that de-identification could mitigate some concerns. However, since explicit consent (from students or parents for younger students) for open data sharing was not obtained, we are unable to release the dataset publicly.

Minor issues:

p 1. lines 1-3: 'Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school student's civic engagement for climate protection' students' instead of student's

p. 6 line 205: '1.4' is repeated

p. 10 lines 286 - 87 'Ratings of three and lower indicate a somewhat negative response and ratings of four and higher indicate a somewhat positive response' - I think this explanation is not necessary and can be omitted

p.11 lines 300-305 The explanation of how 'my environment' is explained would fit better in a footnote

p. 21 lines 586 - 87 the individual 'psychosocial variables' should be specifically mentioned again here

Environmental Psychology Open

Response: Apologies for this oversight, we have amended all minor issues mentioned.

In summary, the instrument presented in the manuscript still needs to be further developed and validated. However, interested researchers, not only from the field of environmental psychology, but also from empirical educational research on ESD, should have access to the instrument. I therefore support the publication of this paper, provided that the aforementioned revisions (especially a detailed description of the test development process and a discussion of limitations) are implemented.

Response: Thank you for taking the time to review our manuscript!

Environmental Psychology Open

Review Round 2

Review A

I thank the authors for submitting a revised version of the paper "Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school student's civic engagement for climate protection", which I reviewed earlier. The authors have made substantial changes both regarding the theoretical framing and the statistical analyses, and responded to all point raised in detail. I appreciate their openness to use alternative statistical approaches and believe that the manuscript can be published after the authors have addressed some additional concerns:

Comment 1: I am still missing a clear explanation of what the scale is measuring. The authors have nicely adjusted the definition of what constitutes CECP and how it differs from other environmentally relevant behavior, but based on the presented analyses (EFA, Rasch-type analyses etc.) they seem to assume that the scale measures a reflective and not a formative construct. So, there should be a latent construct (as also mentioned on page 6, line 176ff) underlying the behaviors of the scale. For me, it would be sufficient, if the authors assume and state that the scale measures a person's inclination or motivation to civically engage for climate protection. But I don't concur with the current notion that the scale reflects CECP itself (e.g., page 7, lines 248f), as this would assume formative or impact-reflecting measurement (see also, Lange, 2024, 2025).

The same applies to the "individual consumption" (and "social influence") scale. The authors present a CFA for the scale, so I wonder what they assume the items reflect. For example, is the scale supposed to reflect a person's carbon footprint? Or their inclination to generally reduce individual consumption?

Comment 2: I have some issues with Table B9 in the Appendix: First, the authors should explain why they chose the measurement model they present for each construct. I assume that they compared model fit between more conservative and more lenient model (e.g., tau-equivalent model vs. essentially tau-equivalent model) by means of a Chi-square test or by comparing BIC or some other fit statistic. The reasoning for choosing each model should be explained, even though I don't expect the authors to present all tested models and their comparisons (maybe just describe the general approach to model selection).

Second, I struggle with understanding the presented degrees of freedom in the table. It would be easier to follow the table, if the authors explained their identification method (e.g., marker method would be the default in lavaan). Currently, the dfs do not always add up. For example, without explanation of the identification method the measurement model for ingroup norms would be just-identified and not over-identified by one parameter as presented in the table (14 known parameters vs. 4 factor loadings, 4 intercepts, 4 error variances, 1 factor variance, and 1 error covariance). I also do not understand how the authors got the dfs for self-efficacy and ingroup identification: essential tau-equivalence refers to the case when item loadings for each item are fixed/equal, while (in contrast to a tau-equivalent model) the item intercepts are allowed to vary across items. Hence, footnote 2 of this table surprised me given that this is already the case in an essentially tau-equivalent model. The authors should a) be more transparent about the model specification and b) check the presented values in the table which can also be computed manually.

Lastly, the abbreviation of CFI in the note does not match the abbreviation in the table (CIF vs CFI) and the p-, RMSEA-, and CFI-values in the table should not have a zero before the decimals as their range is 0-1.

Comment 3: Again, I appreciate the utilization of IRT methods. Based on the description in 2.4, I think that a more straightforward way to testing unidimensionality of the three scales (as described on page 14f) would be a computing a measurement model, in which all items from all scales reflect one latent factor. A Rasch factor analysis could then be performed based on the residuals from this model to investigate whether relevant common variance of items cannot

be captured by this one-dimensional model. I would encourage the authors to test unidimensionality in this way and not how they currently do it, as their current approach might lead on an underestimation of common variance between the three scales.

Based on the presented results you did do exactly that so maybe you can reformulate the method section to make this clearer. It would be helpful to also present fit indices (e.g., item fit and person fit) for this joint model calibration. One component exceeding 2 with 16 items in the model could still be interpreted as evidence for unidimensionality. Item and/or person misfit would be stronger indicators of violations of unidimensionality. Also, the authors should present the items which formed the factor in the Rasch factor analysis. Do they belong to the same scale? If not, this would not indicate separability of the three scales. Accordingly, I currently to not concur with the authors statement that the data supported discriminant validity. This cannot be derived from the presented findings. I understand, that unidimensionality of the three scales would not fit the current argumentation of this manuscript, but it certainly would be an interesting finding, which would also have some theoretical grounding (Kaiser, 2021). Of course, if the model combining all three scales indicates misfit, this would support your current notion of discriminant validity.

On page 18, it is not clear to me whether the Rasch factor analysis is done for the CECP scale alone or the items from the three scales combined.

Turning to the description of the Rasch rating scale model in the results, the infit mean square values are quite extreme, either indicating overfit or misfit. This might be a result of your model choice given that rating scale models impose a restriction on item thresholds (distance between thresholds equal across items). A model that does not have this restriction, such as a Partial Credit Model, might not lead to these more extreme fit values. This could be either discussed or you could calculate the corresponding model which can be easily done with tam (should be the default for polytomous data in tam). In general, this description could also be presented in 3.3 given that it is also a test of scale dimensionality.

Minor remarks:

Abstract: "All items had right-skewed distributions." -> currently, it reads like this applies to all items assessed in the study and not only the CECP items. Is this the case?

Page 4, lines 118ff: The example the authors provide here does not really match the following sentence in which they talk about transformation of social structure. I would argue that building cycle paths does not transform social structures, but rather infrastructure. Maybe the authors can either provide a different example that better fits the following sentence or adjust the wording of the sentence.

Page 4, lines 132ff: I like that the authors introduce the role of education here, but I had some difficulty in following the structure of this paragraph. It might be helpful to move the sentence on education up, directly after first mentioning transformative learning in line 130. Also, I would move the reference to Figure 1 to the beginning of the paragraph, maybe after the first sentence in lines 110ff.

Table 1: The n for missings on the country of growing up is missing.

Page 14, line 382: "on a 6- point Likert" instead of "6-point"

Page 14, line 404: Maybe use polytomous Rasch model instead of Rasch measurement in bracktes.

Environmental Psychology Open

Table 3: It might be helpful to report that the factor loadings are extracted from an EFA. I would also rename the column to Item response theory to be consistent with the CTT column.

Page 18, lines 521ff: This is a nice finding and a nice analysis! Maybe, the authors can clarify the type of Rasch model which was used here (I assume the Rasch rating scale model) to be more transparent. Also, I would suggest to present the Rasch-related analyses together to increase readability.

Table 5, "social influence" row: [57,.69] instead of [.57,.69].

Page 20, lines 576ff: This paragraph could be moved down to the limitations.

Page 21, lines 631ff: In this paragraph, the authors could nicely discuss the Differential Item Functioning analysis they did, which supports the applicability of this scale for various age groups.

Thank you again for the opportunity to read this manuscript. It has significantly improved from the previous submission and I hope that my comments help to further improve it before its publication.

References

Kaiser, F. G. (2021). Climate change mitigation within the Campbell paradigm: doing the right thing for a reason and against all odds. *Current Opinion in Behavioral Sciences*, *42*, 70-75. https://doi.org/10.1016/j.cobeha.2021.03.024

Lange, F. (2024). What is measured in pro-environmental behavior research?. *Journal of Environmental Psychology*, 102381. https://doi.org/10.1016/j.jenvp.2024.102381

Lange, F. (2025). Measurement approaches in climate action research. *Current Opinion in Behavioral Sciences*, *63*, 101510. https://doi.org/10.1016/j.cobeha.2025.101510

Review B

This review refers to the revised version of the manuscript "Measuring socio-ecological handprints: Psychometric evaluation of a scale assessing school students' civic engagement for climate protection." which presents the development and psychometric testing of a scale intended to assess students' civic engagement for climate protection.

In my initial review, I supported the publication of the paper, provided that several concerns (especially a detailed description of the test development process and a discussion of limitations) are addressed in a revised version.

The revised manuscript shows that the authors have responded carefully to each of these requests. Before concluding, I would like to offer two comments for the authors' consideration:

1. With regard to the term "nomological validity", I would have expected a system of theoretically derived hypotheses that would then be tested, for example, using structural equation modeling. The current approach, using individual bivariate correlations, may not fully reflect the theoretical network implied. Therefore, the authors might consider whether "convergent validity" would be a more precise term.

Environmental Psychology Open

2. According to the criterion proposed by D'Haenens (2010; ICC > 0.08), the ICC of .087 indeed indicates the appropriateness of a multilevel analytical approach. However, given that the value is close to the threshold, no substantial bias is to be expected.

From my perspective, the revised version now meets the standards required for publication. I recognise the efforts of the authors to improve the transparency with regard to the test development process and the discussion of the study's results and limitations. The instrument can be valuable not only for researchers in environmental psychology, but also for those working in the broader field of education for sustainable development. I support the publication of the manuscript in its current form and thank the authors for their constructive revision.

References

D'Haenens, E., Van Damme, J., & Onghena, P. (2010). Multilevel exploratory factor analysis: Illustrating its surplus value in educational effectiveness research. School Effectiveness and School Improvement, 21(2), 209 235. https://doi.org/10.1080/09243450903581218

Authors' Response to Reviewer 1:

Comment 1: I am still missing a clear explanation of what the scale is measuring. The authors have nicely adjusted the definition of what constitutes CECP and how it differs from other environmentally relevant behavior, but based on the presented analyses (EFA, Rasch-type analyses etc.) they seem to assume that the scale measures a reflective and not a formative construct. So, there should be a latent construct (as also mentioned on page 6, line 176ff) underlying the behaviors of the scale. For me, it would be sufficient, if the authors assume and state that the scale measures a person's inclination or motivation to civically engage for climate protection. But I don't concur with the current notion that the scale reflects CECP itself (e.g., page 7, lines 248f), as this would assume formative or impact-reflecting measurement (see also, Lange, 2024, 2025).

The same applies to the "individual consumption" (and "social influence") scale. The authors present a CFA for the scale, so I wonder what they assume the items reflect. For example, is the scale supposed to reflect a person's carbon footprint? Or their inclination to generally reduce individual consumption?

Response: We have revised section 1.5 to describe what the scale measures more accurately in reference to the suggested literature:

p.7, l. 273: Our goal was to develop a scale that measures between-person and within-person differences over time in a construct reflecting the propensity to show naturally occurring civic engagement for climate protection in an adolescent's personal environment. According to Lange (2025), the scale measures the propensity to show civic engagement for climate protection (i.e., person property). For the sake of readability, we refer to this construct as civic engagement for climate protection throughout the paper.

Comment 2: I have some issues with Table B9 in the Appendix: First, the authors should explain why they chose the measurement model they present for each construct. I assume that they compared model fit between more conservative and more lenient model (e.g., tau-equivalent model vs. essentially tau-equivalent model) by means of a Chi-square test or by comparing BIC or some other fit statistic. The reasoning for choosing each

Environmental Psychology Open

model should be explained, even though I don't expect the authors to present all tested models and their comparisons (maybe just describe the general approach to model selection).

Second, I struggle with understanding the presented degrees of freedom in the table. It would be easier to follow the table, if the authors explained their identification method (e.g., marker method would be the default in lavaan). Currently, the dfs do not always add up. For example, without explanation of the identification method the measurement model for ingroup norms would be just-identified and not over-identified by one parameter as presented in the table (14 known parameters vs. 4 factor loadings, 4 intercepts, 4 error variances, 1 factor variance, and 1 error covariance). I also do not understand how the authors got the dfs for self-efficacy and ingroup identification: essential tau-equivalence refers to the case when item loadings for each item are fixed/equal, while (in contrast to a tau-equivalent model) the item intercepts are allowed to vary across items. Hence, footnote 2 of this table surprised me given that this is already the case in an essentially tau-equivalent model. The authors should a) be more transparent about the model specification and b) check the presented values in the table which can also be computed manually.

Lastly, the abbreviation of CFI in the note does not match the abbreviation in the table (CIF vs CFI) and the p-, RMSEA-, and CFI-values in the table should not have a zero before the decimals as their range is 0-1.

Response:

We have explained our approach to model selection and the identification method in the table notes:

The loading of the first item on each factor was constrained to be one. Measurement models were selected based on Chi-square difference testing between nested models. More parsimonious models were retained when no significant deterioration in model fit was observed.

Apologies for our oversight regarding the measurement model of the self-efficacy scale, we have corrected this to a tau-congeneric model. We have also corrected the abbreviation of CFI and removed the zeros.

Comment 3: Again, I appreciate the utilization of IRT methods. Based on the description in 2.4, I think that a more straightforward way to testing unidimensionality of the three scales (as described on page 14f) would be a computing a measurement model, in which all items from all scales reflect one latent factor. A Rasch factor analysis could then be performed based on the residuals from this model to investigate whether relevant common variance of items cannot be captured by this one-dimensional model. I would encourage the authors to test unidimensionality in this way and not how they currently do it, as their current approach might lead on an underestimation of common variance between the three scales.

Based on the presented results you did do exactly that so maybe you can reformulate the method section to make this clearer. It would be helpful to also present fit indices (e.g., item fit and person fit) for this joint model calibration. One component exceeding 2 with 16 items in the model could still be interpreted as evidence for unidimensionality. Item and/or person misfit would be stronger indicators of violations of unidimensionality. Also, the authors should present the items which formed the factor in the Rasch factor analysis. Do they belong to the same scale? If not, this would not indicate separability of the three scales. Accordingly, I currently to not concur with the authors statement that the data supported discriminant validity. This cannot be derived from the presented findings. I understand, that unidimensionality of the three scales would not fit the current argumentation of this manuscript, but it certainly would be an interesting finding, which would also have some theoretical grounding (Kaiser, 2021). Of course, if the model combining all three scales indicates misfit, this would support your current notion of discriminant validity.

Environmental Psychology Open

On page 18, it is not clear to me whether the Rasch factor analysis is done for the CECP scale alone or the items from the three scales combined.

Turning to the description of the Rasch rating scale model in the results, the infit mean square values are quite extreme, either indicating overfit or misfit. This might be a result of your model choice given that rating scale models impose a restriction on item thresholds (distance between thresholds equal across items). A model that does not have this restriction, such as a Partial Credit Model, might not lead to these more extreme fit values. This could be either discussed or you could calculate the corresponding model which can be easily done with tam (should be the default for polytomous data in tam). In general, this description could also be presented in 3.3 given that it is also a test of scale dimensionality.

Response:

Thank you for your comments. We have revised section 2.4 to describe the analysis more clearly:

p. 14, l. 443: Further, we computed a Rasch rating scale measurement model in which all items measuring climate protection behaviors (i.e., civic engagement, individual consumption, and social influence) reflect one latent factor. Principal component analysis was performed based on the residuals from this model to test one-dimensionality.

We have added more statistical information about the potential one-factor Rasch rating scale model including all climate protection behaviors in supplemental material C1, on which we reported in the Result section.

p. 19, l.577: Principal component analysis on Rasch residuals was performed on a Rasch rating scale model of all items included in the civic engagement, individual consumption and social influence scales. The first component of the principal component analysis on Rasch residuals had an eigenvalue of 2.11, which exceeds the upper threshold of 2.0 suggested to meet the assumption of one-dimensionality (Linacre, 2022). The principal component analysis suggests a clustering of residuals of the items of the civic engagement scale. In this joint model, some items were outside the suggested range of 0.6 to 1.4 for MNSQ fit statistics (Bond & Fox, 2015) (see supplementary material C1).

In response to your concerns regarding discriminant validity we have revised the manuscript and omitted statements on discriminant validity.

We have reworded 3.3 to clarify that our analysis was done for the model with the eight civic engagement items:

p. 18, l., 529: In addition, one-dimensionality of the scale was assessed by performing principal component analysis on Rasch residuals on the Rasch rating scale model with the eight civic engagement items.

We have added a section discussing the model choice:

p. 23, l. 737: The rating scale model was applied to the data for its robustness against a small amount of observations in some response categories (Linacre, 2000), however, its restrictions on item thresholds might have led to extreme values in the fit indices.

Minor remarks:

Abstract: "All items had right-skewed distributions." -> currently, it reads like this applies to all items assessed in the

Environmental Psychology Open

study and not only the CECP items. Is this the case?

Response: We have revised this sentence to clarify that this refers to the items of the civic engagement for climate protection scale.

Page 4, lines 118ff: The example the authors provide here does not really match the following sentence in which they talk about transformation of social structure. I would argue that building cycle paths does not transform social structures, but rather infrastructure. Maybe the authors can either provide a different example that better fits the following sentence or adjust the wording of the sentence.

Response: We have revised the wording to create a better fit with the example.

p. 4, l.146**:** The impact that is generated by the transformation of systemic conditions towards the mitigation of climate change can be significantly higher than the impact of individual behavior change and is crucial for a socio-ecological transformation towards climate protection (Chater & Loewenstein, 2022).

Page 4, lines 132ff: I like that the authors introduce the role of education here, but I had some difficulty in following the structure of this paragraph. It might be helpful to move the sentence on education up, directly after first mentioning transformative learning in line 130. Also, I would move the reference to Figure 1 to the beginning of the paragraph, maybe after the first sentence in lines 110ff.

Response: We have re-arranged this paragraph and implemented the suggested changes.

Table 1: The n for missings on the country of growing up is missing.

Response: Apologies for this oversight, we have corrected this.

Page 14, line 382: "on a 6- point Likert" instead of "6-point"

Response: Apologies for this oversight, we have corrected this.

Page 14, line 404: Maybe use polytomous Rasch model instead of Rasch measurement in bracktes.

Response: We have changed this as suggested.

Table 3: It might be helpful to report that the factor loadings are extracted from an EFA. I would also rename the column to Item response theory to be consistent with the CTT column.

Response: We have made the suggested changes to table 3.

Page 18, lines 521ff: This is a nice finding and a nice analysis! Maybe, the authors can clarify the type of Rasch model which was used here (I assume the Rasch rating scale model) to be more transparent. Also, I would suggest to present the Rasch-related analyses together to increase readability.

Environmental Psychology Open

Response: Thank you for this suggestion, we have incorporated that we used the Rasch rating scale model. We have opted to include the Rasch-related analysis in the respective results sections, structured by the research questions.

Table 5, "social influence" row: [57,.69] instead of [.57,.69].

Response: Apologies for this oversight, we have corrected this.

Page 20, lines 576ff: This paragraph could be moved down to the limitations.

Response: We have moved this paragraph to the limitations.

Page 21, lines 631ff: In this paragraph, the authors could nicely discuss the Differential Item Functioning analysis they did, which supports the applicability of this scale for various age groups.

Response: We have added this, which reads:

p. 22, l. 670: Our analysis indicates that the scale is applicable for adolescents of different ages.

Thank you again for the opportunity to read this manuscript. It has significantly improved from the previous submission and I hope that my comments help to further improve it before its publication.
Response: Thank you, we are grateful for the time and care you invested in reviewing our manuscript.

Authors' Response to Reviewer 2:

We would like to sincerely thank the reviewer for their constructive feedback on our revised manuscript. We address the comments below:

1. We agree with your point and have removed the term "nomological validity".
2. Thank you for pointing this out. We have updated the manuscript to appropriately contextualize the ICC result, which reads:

p. 18, l. 541**:** Hence, no substantial bias due to the multi-level structure of the data is expected.