



# (How Much) Do Temporal Social Comparisons Matter?

A Replication of Study 1a of Reh, Tröster, and Van Quaquebeke (2018)

Roman Briker<sup>1</sup> and Frank Walter<sup>2</sup> 

<sup>1</sup>Department of Organization, Strategy, and Entrepreneurship, Maastricht University, Maastricht, The Netherlands

<sup>2</sup>Department of Organization and Human Resource Management, Justus Liebig University, Giessen, Germany

**Abstract:** Moving beyond static perspectives in social comparison theory, Reh and colleagues (2018) provided initial evidence for the relevance of “temporal social comparisons” (i.e., comparing one’s own with others’ past development over time on a salient dimension). Although this research has received wide attention, the study illustrating the authors’ basic rationale (Study 1a) suffered from a small sample size, and its results did not reach conventional significance levels. Thus, we provide a direct, preregistered, and high-powered replication of this study. Our results corroborate the original conclusions, indicating that unfavorable temporal social comparisons evoke social undermining in more (but not less) competitive contexts. These findings reiterate the importance of a dynamic, temporal perspective for a complete understanding of social comparison processes.

**Keywords:** social comparison, social undermining, replication

Social comparison is widely studied in social and applied psychology, with individuals comparing themselves to others along various dimensions (e.g., performance, skills, or characteristics) to self-evaluate their social standing (Baldwin & Mussweiler, 2018; Festinger, 1954). Scholars have demonstrated the profound consequences associated with such comparisons (see Gerber et al., 2018). In interpersonal interactions, in particular, research has shown that unfavorable social comparisons (i.e., perceiving others as superior) can lead individuals to undermine or harm others (e.g., Duffy et al., 2012; Lam et al., 2011).

Traditionally, this research has examined social comparisons at a single time point, largely ignoring the possibility that such processes may involve a pronounced temporal component. This is in contrast to the literature on *self-comparisons*, which suggests that individuals consider, for example, their own past and anticipated future performance (or skills) when making self-assessments (Van Yperen & Leander, 2014; Wilson & Ross, 2000). More recently, however, Reh et al. (2018) have proposed a more dynamic conceptualization of *social comparisons*. Beyond comparing one’s current standing with salient others, these authors have emphasized the role of *temporal social comparisons*, such that individuals contrast their development on

relevant dimensions over time with other persons’ respective development.

Reh et al. argued, in particular, that individuals compare the trajectories of their own and others’ past performance. If this temporal social comparison is unfavorable (i.e., a target’s performance has developed more positively over time than one’s own), perceptions of future status threat are suggested to arise, triggering social undermining behavior toward the comparison target to counter such threats (i.e., by elevating one’s own status at the target’s expense). Moreover, Reh et al. hypothesized these consequences to be particularly pronounced in competitive situations that heighten individuals’ sensitivity to status threats. The authors provided empirical support for this rationale across various studies, including both experimental and field designs.

In doing so, Reh et al.’s investigation has opened up new directions for the social comparison literature. Moving beyond traditional, more static approaches, their findings highlight the relevance of a dynamic perspective for a complete understanding of social comparison processes, taking into account both focal individuals’ and comparison targets’ past developments. Moreover, by illustrating competition as a key boundary condition, Reh et al. explain why such

temporal social comparisons are more relevant in some situations than others. Despite its relatively recent publication, Reh et al.'s research has received considerable attention from social and organizational psychologists, and it has informed social comparison theory in important ways (e.g., DeGagne & Busseri, 2021; Fasbender & Gerpott, 2021).<sup>1</sup>

## The Current Research

Consistent with prior replication research (e.g., Mayiwar & Lai, 2019; Tybur et al., 2020), we aim to replicate the foundational experimental test of Reh et al.'s core hypothesis (i.e., the interactive effect of temporal social comparison and competition on social undermining), as reflected in their Study 1a. This is, to our knowledge, the first replication of this experiment. Reh et al. have also conducted an additional, very similar experiment (Study 1b), a vignette study (Study 2), and a field investigation (Study 3). Importantly, however, Study 1a's dependent variable measure taps more closely into the authors' basic theoretical reasoning than Study 1b. This is because it assesses social undermining behavior that may advance a participant's outcomes at the target's expense. Moreover, as Reh et al. noted, Study 2's vignette approach suffers from limited realism and potential demand effects, and Study 3's survey design does not allow for causal conclusions. Hence, we believe replicating Study 1a is particularly useful for scrutinizing Reh et al.'s core theoretical argumentation and testing the robustness of their central finding. Moreover, although providing preliminary support for the hypothesized interaction effect, Study 1a's sample size is very small ( $N = 90$ ), and the respective coefficient estimate was only significant at the 10% level ( $p = .086$ ). Hence, a higher-powered replication may clarify this somewhat ambiguous finding and provide further confidence in its viability.

Consequently, this study provides a direct, preregistered, and high-powered replication of Reh et al.'s Study 1a. Mirroring the original procedures, manipulations, and measures, we revisited the hypothesis that unfavorable temporal social comparisons enhance individuals' social undermining in competitive (but not non-competitive) situations. Scholars have cast such replication as critical for establishing a solid, reliable evidence base – and the replication crisis that has recently plagued psychology and other fields underlines this relevance (Shrout & Rodgers, 2018). Moreover, given the recency of Reh et al.'s ideas and insights, we believe empirical scrutiny of their key findings is particularly timely, enabling researchers to more confidently decide whether a temporal, dynamic perspective is

useful for social comparison theory and whether investing future resources into studying this new view is worthwhile.

## Method

This investigation was part of a larger project entitled SCORE (Systematizing Confidence in Open Research and Evidence) that aims to assess the replicability of findings in the social and behavioral sciences. This project involved more than one thousand contributing researchers, from which small teams (including the present article's authors) were tasked with replicating a specific empirical finding, as chosen by SCORE's core project team (for more information on the overall project, see SCORE Collaboration, 2021).

The present research directly replicated Reh et al.'s Study 1a, following the original study as closely as possible. To facilitate this, two independent reviewers and one editor within the overall SCORE project vetted our design, and two of the original study's authors provided further comments to ensure that all procedures, manipulations, and measures were fully equivalent. In fact, the only material difference between the original study and the current procedures is that we excluded individuals that had already participated in the original study. We received IRB approval for the final research design (BRANY protocol no. 20-026-757) and preregistered it; all preregistration documents and further information (including study materials, data, and analyses) are available online (see Briker & Walter, 2021). In the following, we report how we determined our sample size, data exclusions, and manipulations and measures.

## Power Analysis, Participants, and Design

We aimed to achieve at least 90% power to detect 75% of the original study's effect size for the key hypothesis test. Power analyses using R revealed that a usable sample size of 573 participants was necessary to meet this criterion. Mirroring Reh et al., we recruited US-based participants from Amazon's MTurk, and we limited participation to individuals who had completed at least 50 MTurk tasks with a minimum approval rate of 95%. In addition, we used comprehension checks to ensure that participants had understood basic task instructions and procedures, providing additional clarification if these checks were incorrectly answered. Finally, we excluded participants who doubted the study's realism (again following Reh et al., 2018) or indicated that they had previously participated in the same or a highly similar study.

To reach the desired sample size despite these exclusion criteria, we targeted a substantively greater number of

<sup>1</sup> Although only published in 2018, Reh et al. have received 61 citations in Google Scholar by July 29, 2021.

individuals than required, and we obtained data from 898 participants. We excluded 134 of these participants because they expressed doubts about study realism (i.e., 15% of the initial sample, compared to 17% in the original study), and 190 participants were excluded because they indicated previous study participation. Hence, our final sample comprised 574 participants (43% female;  $M_{\text{age}} = 38.33$ ,  $SD_{\text{age}} = 10.91$ ).<sup>2</sup> As in the original study, participants were randomly assigned to an experimental condition in a  $2 \times 2$  mixed design, manipulating temporal social comparisons (favorable vs. unfavorable) as a within-subjects factor and competition (high vs. low) as a between-subjects factor.

## Procedure

Following Reh et al., we invited participants to a study on “Intellectual Performance in the Presence of a Co-actor” for a fixed compensation of \$1.00 and a chance to win an additional bonus of \$1.00. After providing informed consent, they signed into an alleged chatroom (using a nickname), designed to underscore the impression that others simultaneously took part in the study and could later be matched with a focal participant. After less than 1 min, participants were automatically directed to the main study.<sup>3</sup>

The participants performed a series of verbal ability tests (i.e., anagram tasks). To increase participants’ engagement and enhance the salience of task performance as a social comparison standard, we emphasized that this task type measures analytical reasoning skills important for many life domains. Moreover, we informed the participants that performance on such tasks could develop over time, ascertaining the plausibility of the temporal performance trajectories subsequently presented. As in Reh et al., participants started by completing six practice anagrams, and we subsequently provided the respective solutions. In a second step, participants performed five 1-minute rounds comprising 20 anagrams each, with the goal of solving as many anagrams as possible. We framed this study phase as “training rounds,” designed to increase participants’ task familiarity before the final study part (which did not take place).<sup>4</sup> Participants were told that within this (alleged) final part, they would be matched with a co-participant to again complete anagram tasks, with the chance of earning a \$1.00 bonus.

As outlined below, the competition manipulation was embedded within the task instructions and reiterated directly after the alleged training rounds (to ensure its salience). Moreover, participants also received the temporal social comparison manipulation after the training rounds, designed as bogus performance feedback. Finally, after this feedback, we measured the dependent variable and informed the participants that the study was over at this point. Participants then answered questions about study realism and demographics, were debriefed and thanked, and received \$2.00 compensation (regardless of their performance).

## Manipulations

### Temporal Social Comparison

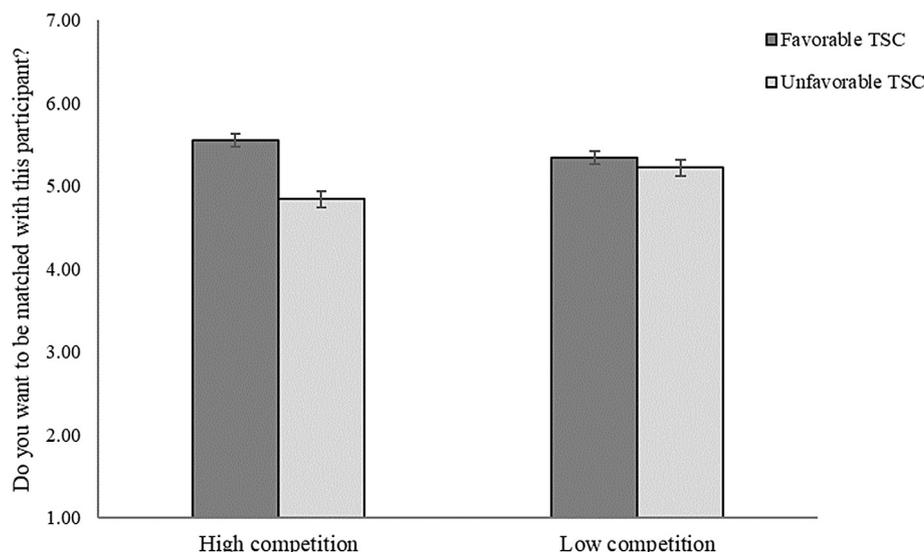
Consistent with Reh et al., we manipulated temporal social comparison as a within-subjects factor, subsequently showing two bogus feedback graphs to participants after the training rounds (see Electronic Supplementary Material, ESM 1, Figures E1 and E2). These graphs compared a participant’s alleged performance trajectory during the five training rounds with the respective trajectories of two potential co-participants for the final study part. To ensure that this feedback was credible, we emphasized that it (a) considered the total number of correct solutions, as well as the length and difficulty of the anagrams and (b), was shown relative to previous participants rather than in absolute terms.

These graphs depicted participants’ own performance as relatively stable, with minor fluctuations over time to increase realism. In the favorable temporal social comparison condition, this performance development was compared with a co-participant whose performance slightly decreased over the five training rounds. In the unfavorable temporal social comparison condition, by contrast, the co-participant’s performance strongly increased. Importantly, the co-participant’s performance in the fifth training round was depicted as similar across experimental conditions (i.e., slightly below the focal participant). Hence, the conditions differed in terms of *temporal* social comparison (with different performance trajectories) but not *current* social comparison (with identical performance levels in the last round). We randomized the order in which these graphs were presented.

<sup>2</sup> As preregistered, we had initially estimated that collecting data from 688 participants would be sufficient to achieve our target of 573 usable responses. However, we had to exclude more individuals than expected due to them having indicated prior study participation. Hence, reaching our targeted usable sample size required collecting data from 898 participants. Importantly, all data analyses were performed only after we had reached the targeted sample size. Of the excluded participants, 148 were in the high competition condition and 176 in the low competition condition. Competition condition was not significantly related to exclusion ( $p = .073$ ). In the final sample, 298 participants were in the high competition condition and 276 in the low competition condition (temporal social comparison was manipulated as a within-subject factor).

<sup>3</sup> This relatively complex procedure, along with the open-ended format of the experimental task, also reduced the risk of automated responding (e.g., bots; Aguinis et al., 2021).

<sup>4</sup> This framing is consistent with the original study (Reh, personal communication, July 14, 2020).



**Figure 1.** Mean ratings of willingness to be matched with a co-participant as a function of temporal social comparison and competition. Error bars denote standard errors.

### Competition

We manipulated competition as a between-subjects factor when providing task instructions. In the high competition condition, we informed the participants that they would only receive the \$1.00 bonus if they outperformed their co-participant in the final study part. In the low competition condition, participants were told that they needed to exceed a given performance threshold in the final part to earn the bonus, independent of their co-participant's performance. We reiterated this manipulation through a brief reminder directly before presenting the feedback graphs.

### Dependent Variable

Again, following Reh et al., we measured social undermining through participants' (un-)willingness to be matched with a co-participant. Directly after seeing each of the performance feedback graphs, participants rated their willingness to be matched with this potential co-participant on a scale from 1 (= *I do not want to be matched with this participant at all*) to 7 (= *I very much want to be matched with this participant*). Importantly, participants were told that a co-participant not matched with them would have to wait for another possible partner, missing the chance to immediately win the bonus. Hence, lower scores on this measure reflect greater social undermining of a co-participant.

## Results

### Preregistered Analyses

As in the original study, we used multilevel ordered logistic regression analyses (Sommet & Morselli, 2017) to test the joint effects of temporal social comparison (within-subjects)

and competition (between-subjects) on social undermining, because the outcome variable was measured on an ordinal scale (Jamieson, 2004). Mirroring Reh et al., these analyses revealed no simple effects for either temporal social comparison ( $\log\text{-odds} = -0.143$ ,  $SE = 0.148$ ,  $p = .335$ , 95% CI  $[-0.433; 0.147]$ ; Cohen's  $d = 0.07$ ) or competition ( $\log\text{-odds} = 0.313$ ,  $SE = 0.172$ ,  $p = .069$ , 95% CI  $[-0.025; 0.651]$ ;  $d = 0.17$ ).

Importantly, however, our analyses demonstrated a significant Temporal Social Comparison  $\times$  Competition interaction ( $\log\text{-odds} = -0.736$ ,  $SE = 0.227$ ,  $p = .001$ , 95% CI  $[-1.181; -0.290]$ ;  $d = 0.41$ ). Planned contrasts (see Figure 1) showed that, with high competition, temporal social comparison had a significant effect on social undermining ( $\log\text{-odds} = -0.879$ ,  $SE = 0.167$ ,  $p < .001$ , 95% CI  $[-1.206; -0.551]$ ;  $d = 0.48$ ), such that participants were less willing to be matched with a co-participant in the unfavorable rather than favorable comparison condition. With low competition, by contrast, the effect of temporal social comparison on undermining was not significant ( $\log\text{-odds} = 0.586$ ,  $SE = 0.344$ ,  $p = .088$ , 95% CI  $[-0.088; 1.260]$ ;  $d = 0.32$ ). Overall, these results replicate Reh et al.'s pattern of findings (see Table E1 in ESM 1, for a side-by-side comparison).

### Exploratory Analyses

To explore our findings' robustness, we repeated our analyses (a) without excluding participants that indicated to have participated in the original study and (b) without excluding any participants. In both cases, the results were nearly identical to the ones reported before (see ESM 1, Table E2). Moreover, we repeated the hypotheses test

using a mixed-design analysis of variance (ANOVA). These alternative analyses again revealed a significant Temporal Social Comparison  $\times$  Competition interaction effect on social undermining,  $F(1, 572) = 11.45, p = .001, \eta^2 = .02$ , and the interaction pattern was equivalent to the one previously reported.

## Discussion

Extending traditional, static approaches in social comparison theory, Reh et al. (2018) have introduced a new, more dynamic perspective to this literature. They illustrated that people draw from their own and others' performance *developments* over time when making social comparisons. Specifically, the authors showed that unfavorable temporal social comparisons (i.e., a comparison target's performance developing more positively than one's own) increased social undermining under conditions of high (but not low) competition, largely irrespective of a target's current performance level.

To examine the robustness of this finding, the present research offers a preregistered, independent, and high-powered replication of Reh et al.'s foundational study (i.e., Study 1a), mirroring the original study's design. Importantly, our core results align with Reh et al.'s findings, thus corroborating their initial conclusions – although the effect size for the Temporal Social Comparison  $\times$  Competition interaction was slightly smaller than in the original study (i.e., Cohen's  $d = 0.41$  vs.  $0.52$ ; see ESM 1, Table E1). Our study, therefore, increases confidence in the role of unfavorable temporal social comparisons for social undermining, and it reiterates that such comparisons are particularly important (and potentially damaging) in more rather than less competitive situations. More generally, our findings underscore the viability of a temporal, dynamic perspective on social comparison processes, illustrating that future theory and research may substantively broaden our understanding of such processes by systematically considering individuals' past developments on relevant comparison dimensions. Finally, our results (in conjunction with Reh et al.'s findings) show that this temporal perspective may be particularly relevant for social comparisons in specific contexts, for example, in organizations with a highly competitive work environment.

Finally, some limitations of our investigation deserve mention. As noted earlier, we focused on Reh et al.'s Study 1a to investigate their core theoretical rationale. Given our study's supportive evidence in this regard, we believe future research may benefit from additional replication efforts pertaining to Reh et al.'s subsequent studies (e.g., on the mechanisms underlying the effects uncovered

in Study 1a). Moreover, our goal was to replicate the original study's procedures as closely as possible. Hence, our findings cannot speak to the generalizability of these findings to alternative contexts (e.g., other cultures) or specific demographic groups, and further research investigating these issues would be valuable.

## Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at <https://doi.org/10.1027/1864-9335/a000458>

**ESM 1.** Figures E1 and E2. Bogus performance graph in the favorable/unfavorable temporal social comparison condition. Table E1. Comparing results between present study and Reh et al. (2018), Study 1a. Table E2. Results with and without exclusion of participants.

## References

- Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, *47*(4), 823–837. <https://doi.org/10.1177/0149206320969787>
- Baldwin, M., & Mussweiler, T. (2018). The culture of social comparison. *Proceedings of the National Academy of Sciences*, *115*(39), E9067–E9074. <https://doi.org/10.1073/pnas.1721555115>
- Briker, R., & Walter, F. (2021). *Study materials for “(How Much) Do temporal social comparisons matter? A replication of Reh, Tröster, and Van Quaquebeke (2018)”*. [https://osf.io/cgyb9/?view\\_only=da83709577b54d82a149f5498e8691c3](https://osf.io/cgyb9/?view_only=da83709577b54d82a149f5498e8691c3)
- DeGagne, B., & Busseri, M. A. (2021). The impact of better- versus worse-than-average comparisons on beliefs about how life satisfaction is unfolding over time, affect, and motivation. *European Journal of Social Psychology*. Advance online publication. <https://doi.org/10.1002/ejsp.2765>
- Duffy, M. K., Scott, K. L., Shaw, J. D., Tepper, B. J., & Aquino, K. (2012). A social context model of envy and social undermining. *Academy of Management Journal*, *55*(3), 643–666. <https://doi.org/10.5465/amj.2009.0804>
- Fasbender, U., & Gerpott, H. F. (2021). Knowledge transfer between younger and older employees: A temporal social comparison model. *Work, Aging and Retirement*. Advance online publication. <https://doi.org/10.1093/workar/waab017>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. <https://doi.org/10.1177/001872675400700202>
- Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177–197. <https://doi.org/10.1037/bul0000127>
- Jamieson, S. (2004). Likert scales: How to (ab)use them? *Medical Education*, *38*(12), 1217–1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>
- Lam, C. K., Van der Vegt, G. S., Walter, F., & Huang, X. (2011). Harming high performers: A social comparison perspective on interpersonal harming in work teams. *Journal of Applied Psychology*, *96*(3), 588–601. <https://doi.org/10.1037/a0021882>

- Mayiwar, L., & Lai, L. (2019). Replication of Study 1 in “Differentiating social and personal power” by Lammers, Stoker, and Stapel (2009). *Social Psychology*, *50*(4), 261–269. <https://doi.org/10.1027/1864-9335/a000388>
- Reh, S., Tröster, C., & Van Quaquebeke, N. (2018). Keeping (future) rivals down: Temporal social comparison predicts coworker social undermining via future status threat and envy. *Journal of Applied Psychology*, *103*(4), 399–415. <https://doi.org/10.1037/apl0000281>
- SCORE Collaboration. (2021). *Systematizing Confidence in Open Research and Evidence (SCORE)*. SocArxiv. <https://doi.org/10.31235/osf.io/46mnb>
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*(1), 203–218. <https://doi.org/10.5334/irsp.90>
- Tybur, J. M., Jones, B. C., DeBruine, L. M., Ackerman, J. M., & Fasolt, V. (2020). Preregistered direct replication of “Sick body, vigilant mind: The biological immune system activates the behavioral immune system”. *Psychological Science*, *31*(11), 1461–1469. <https://doi.org/10.1177/0956797620955209>
- Van Yperen, N. W., & Leander, N. P. (2014). The overpowering effect of social comparison information: On the misalignment between mastery-based goals and self-evaluation criteria. *Personality and Social Psychology Bulletin*, *40*(5), 676–688. <https://doi.org/10.1177/0146167214523475>
- Wilson, A. E., & Ross, M. (2000). The frequency of temporal-self and social comparisons in people’s personal appraisals. *Journal of Personality and Social Psychology*, *78*(5), 928–942. <https://doi.org/10.1037/0022-3514.78.5.928>

## History

Received February 18, 2021

Revision received August 19, 2021

Accepted August 24, 2021

Published online November 12, 2021

## Acknowledgments

We thank the Center for Open Science team involved in the SCORE project, especially Zachary Loomas and Brianna Luis, for their help and guidance in conducting the replication study. We also thank Hannes Gerstel for setting up the virtual chat room. Finally, we thank Jordan Wagge, Bill Chopik, Heather Kappes, Susan Reh, and Niels van Quaquebeke for reviewing the preregistration and providing useful comments.

## Authorship

All authors contributed significantly to the research. Both authors designed the study. Roman Briker prepared the OSF archive for the study, performed data collection, analysis, and drafted the first version of the manuscript. Frank Walter provided feedback throughout data collection and analysis, and critically edited and revised the manuscript.

## Open Data

For the preregistration, see [https://osf.io/h3wbg?view\\_only=37c3cfb10ad14c8eba0686a9bdf9eb1](https://osf.io/h3wbg?view_only=37c3cfb10ad14c8eba0686a9bdf9eb1). For study materials, see [https://osf.io/cgyb9/?view\\_only=da83709577b54d82a149f5498e8691c3](https://osf.io/cgyb9/?view_only=da83709577b54d82a149f5498e8691c3) (Briker & Walter, 2021).

## Funding

Open access publication enabled by Justus Liebig University Giessen.

## ORCID

Frank Walter

 <https://orcid.org/0000-0001-6328-7430>

## Frank Walter

Department of Organization and HRM

Justus Liebig-University Giessen

Licher Straße 62

35394 Giessen

Germany

[frank.walter@wirtschaft.uni-giessen.de](mailto:frank.walter@wirtschaft.uni-giessen.de)