Title

“Am I trying hard or harder than others?”: Gender differences in reciprocal relations between perceived effort, science self-concept, and achievement in chemistry

Journal

Contemporary Educational Psychology, 80

ISSN

0361-476X

Authors

Lee, Hyewon
Yu, Shirley L
Lin, Tzu-Jung
et al.

Publication Date

2025-03-01

DOI

10.1016/j.cedpsych.2025.102349

“Am I trying hard or harder than others?”: Gender differences in reciprocal relations between perceived effort, science self-concept, and achievement in chemistry

Hyewon Lee[1], Shirley L. Yu[2], Tzu-Jung Lin[2], & Minjung Kim[2]

[1] School of Education, University of California Irvine, Irvine, CA, United States

[2] Department of Educational Studies, The Ohio State University, Columbus, OH, United States

Authors’ Note:

For correspondence about this manuscript, please contact Hyewon Lee (hyewl3@uci.edu). This research was partially supported by the Center for Emergent Materials, an NSF MRSEC, under award number DMR-2011876.

Introduction

High attrition rates in the fields of Science, Technology, Engineering, and Mathematics (STEM) during the early college years (Denice, 2021; Lee et al., 2024; NCES, 2021) suggest a need to better understand the motivational processes in the transition from secondary school to university STEM programs. Science self-concept (science SC; Marsh, 1993), defined as the self-perception of one's ability in the domain of science, has been found to be a key precursor of performance and educational/career choices in STEM (Jansen et al., 2015; Nagengast & Marsh, 2012; Nagy et al., 2006; Simpkins et al., 2006; Taskinen et al., 2013; Zhang et al., 2022). Advocates of social-cognitive theories (e.g., Eccles & Wigfield, 2020) and academic self-concept theory (ASC; Marsh, 1993) suggest that ASC is informed by the social context. Therefore, academic transitions are an important social context during which ASC is reshaped by information from the new environment (Gniewosz et al., 2012; von Keyserlingk et al., 2019).

Introductory STEM courses are academic major prerequisites that students usually complete during their transition to college. These courses are often referred to as "weed-out" courses, typically characterized by rigorous coursework, a competitive climate, and high-achieving student populations (Seymour & Hunter, 2019). These courses may challenge the development of students' confidence in STEM (Kim et al., 2022; Meaders et al., 2020) and may lead them to switch majors (Chen, 2013). Moreover, cultural beliefs favoring male superiority (e.g., gender stereotypes such as, "Women are less capable than men in STEM."; Cheryan et al., 2017) and brilliance (e.g., field-specific ability beliefs such as, "Brilliance is required for success in STEM."; Bian et al., 2018) establish gendered barriers - such as discrimination, negative psychological experiences - that disproportionately undermine women's confidence in STEM

(Ertl et al., 2017; Muradoglu et al., 2023; Robnett, 2016). Therefore, examining the gendered developmental processes of science SC within the context of introductory STEM courses is essential for guiding practices that promote gender equity in STEM.

Previous studies have demonstrated the complex relations between effort and ASC (Marsh et al., 2016). Although effort is crucial for succeeding in challenging introductory courses, students may perceive it as inversely related to ability in competitive settings such as STEM (see Jagacinski & Nicholls, 1987 for discussion; Muenks & Miele, 2017). Furthermore, perceived effort may be more critical for shaping science SC of women than men (Smith et al., 2013), as women's high effort may confirm gender stereotypes and contradict the emphasis on brilliance in STEM. Although prior work on the relations between perceived effort and ASC offers valuable insights into the development of ASC, there remains a lack of empirical evidence regarding science SC.

Accordingly, in this study, we examined gender differences in the relations between perceived effort and science SC in undergraduate introductory chemistry courses. We adopted two novel approaches to address this inquiry. First, building on the internal/external frame of reference model (I/E model; Marsh & Craven, 2002), social comparison theory (Festinger, 1954), and Nicholls' work (1978), we differentiated between two types of perceived effort: *criterion effort* (aligned with traditional views of effort) and *comparative effort* (focused on social comparisons of effort). Second, guided by the reciprocal effects model (REM; Marsh et al., 2018), we investigated the roles of perceived effort as both causes and effects of science SC and achievement. Collectively, we investigated gender differences in reciprocal relations between two types of perceived effort (criterion, comparative effort), science SC, and achievement (midterms, final exam) in introductory chemistry.

Social Comparisons in the Formation of ASC

The I/E model (Marsh & Craven, 2002) posits that ASC is shaped through social (comparing one's achievement with that of one's classmates in the same domain), dimensional (comparing one's achievement in a domain with one's own achievement in other domains), and temporal comparisons (comparing one's current achievement with one's own past achievement in the same domain). Although these comparisons occur conjointly, social comparisons have been found to be the most influential for ASC (Wolff et al., 2018; Wolff & Möller, 2022), with relatively higher achievement than peers leading to an enhanced ASC. However, only a few studies have examined these relations for science SC, as existing research has largely focused on math or verbal SC (Jansen et al., 2014, 2015). Therefore, it is crucial to investigate how social comparison processes influence science SC to expand the literature.

Social comparison theory (Festinger, 1954) provides a broader theoretical framework for understanding the social comparison processes underlying ASC. According to the theory, individuals engage in comparisons with others to satisfy a fundamental human need for self-evaluation. In classroom settings, students often rely on grades or other indicators of their relative standing to evaluate their abilities, which, in turn, shape their ASCs associated with various STEM domains (Huguet et al., 2009; Müller-Kalthoff et al., 2017), including science SC (Chiu, 2012; Nagengast & Marsh, 2011, 2012). Social comparisons become particularly prominent in competitive environments (Boissicat et al., 2021; Dijkstra et al., 2008) and during transitional periods due to a changed frame of reference (Festinger, 1954). Thus, the competitive and transitional environments of introductory STEM courses provide an important context for examining how social comparisons relate to science SC.

In STEM, women often have lower science SC than men even when they achieve similar levels of success (Jansen et al., 2014; Liou et al., 2021; Saß & Kampa, 2019; Wilkins, 2004). This discrepancy raises important questions about whether students utilize various ability-related indicators beyond actual performance when self-evaluating their scientific abilities, and whether these indicators contribute to observed gender differences in science SC.

Perceived Effort as a Dimension of Social Comparisons: Criterion Effort vs. Comparative Effort

Complex Relations Between ASC and Effort. Effort is often viewed as a double-edged sword in achievement settings (Covington, 1984; Jagacinski & Nicholls, 1984). Covington (1984) argues that while effort is essential for improving performance, it can also diminish one's ASC when it is accompanied by failures. Jagacinski and Nicholls (1984) suggest that students' perceptions of the relations between effort and ability are influenced by the social context. In academic environments that emphasize competition and social comparisons, students are likely to view ability as negatively related to the effort required to complete tasks (e.g., "Successful students are brilliant and do not have to work hard."). In contrast, in contexts that focus on individual growth or mastery, they are more inclined to see ability as positively related to effort they invest (e.g., "Successful students work hard.").

Researchers have further investigated the complex nature of perceived effort in relation to both the self (e.g., personal values) and tasks (e.g., subjective task difficulty), with each aspect yielding distinct consequences. Specifically, Muenks et al. (2016) found that students can perceive the relations between effort and ability positively when effort is focused on self-improvement (e.g., motivation). In contrast, this relation can be perceived negatively when effort is regarded as a response to task demands (e.g., subjective task difficulty). On a broader level, Inzlicht et al. (2018) provided evidence from cognitive science, neuroscience, and economics that emphasizes the complexity of perceived effort. They noted that individuals generally value outcomes achieved through effort more than those attained with less effort. However, when perceived effort required to succeed becomes excessive in relation to the personal importance of the outcomes, effort may be perceived as a cost (c.f., effort cost; Eccles & Wigfield, 2020) and the outcomes are devalued. These findings indicate the psychological mechanisms that influence whether effort is seen as relevant to an individual's self-concept (e.g., personal ability, values, or goals).

Social Comparisons of Effort. Social comparisons may function as another mechanism through which effort is differentially related to ability. In experimental studies (Jagacinski & Nicholls, 1987; Muenks et al., 2016), undergraduate students were asked to evaluate individuals' abilities after reading vignettes that described the effort they put into tasks. Results consistently showed that students perceived someone as less competent when their high effort was framed in comparison to others ("A person expended more effort than others."; social comparisons of effort). Conversely, when high effort was described without comparisons ("A person expended high levels of effort."), students viewed that person as competent. One study (Smith et al., 2013) directly tested the predictive validity of social comparisons of effort in science SC, providing both empirical and experimental evidence of their negative associations among undergraduate students in STEM.

Taken together, prior evidence highlights the potential importance of social comparisons of effort underlying science SC; however, this area remains largely underexplored. To build on this line of research, we examined the roles of social comparisons of effort in a natural STEM course context by measuring two types of perceived effort. The first construct pertains to students' perceptions of the effort they expend on a task compared to their own typical levels of effort, aligning closely with traditional views of effort in achievement literature (Marsh et al., 2016); we call this *criterion effort.* The second construct involves students' perceptions of the effort they exert on a task in comparison to the effort they believe others expend on equivalent tasks, reflecting social comparisons; we call this *comparative effort*. We established comparative effort in relation to average peers within the same classroom to address the proximity of comparison targets (the local dominance effect; the tendency to make comparisons with local rather than distal targets; Zell & Alicke, 2020).

Reciprocal Relations Between Perceived Effort, ASC, and Achievement

Perceived effort acts as both a precursor and an outcome of ASC (Marsh et al., 2016). Therefore, it is important to examine both types of perceived effort to gain a broader understanding of the development of science SC. The reciprocal effects model (REM; Marsh et al., 2018) posits reciprocal relations between ASC and achievement, whereby prior ASC is linked to future achievement, and prior achievement influences subsequent ASC. REM has been extensively studied across academic domains (e.g., Wu et al., 2021). In STEM, some studies have found reciprocal relations between science SC and achievement among undergraduates (Gibbons & Raker, 2019; Villafane et al., 2016) and high school students (Wang et al., 2008; Zhao et al., 2024).

Motivational processes in REM can be further explained by perceived effort (Marsh et al., 2016). As noted by the theories (Dweck, 2000; Nicholls, 1978), the relations between achievement and ASC can depend on whether outcomes are attributed to perceived effort. Furthermore, perceived effort may mediate the relations between ASC and achievement (Skaalvik & Skaalvik, 2013). To our best knowledge, researchers have not yet explored the roles of perceived effort in REM in the domain of science. Findings from this line of research in math (Marsh et al., 2016; Pinxten et al., 2014; Xu, 2018, 2020) lend support for the correlates of perceived effort in science. For example, Marsh et al. (2016) examined perceived effort in REM in math with German secondary students over four years. Results showed that perceived effort negatively predicted math SC, whereas math achievement positively predicted perceived effort. Furthermore, they found significant interactive effects, such that perceived effort positively predicted subsequent math SC and achievement when prior math SC was high. These findings point to the complex roles of perceived effort in explaining ASC and achievement, underscoring the need for further research to identify the underlying mechanisms. Thus, we addressed this question by investigating two types of perceived effort.

Gender Differences in the Relations Between Perceived Effort and Science SC

The cultural values, norms, and practices in STEM that emphasize masculinity and brilliance often contradict women's scientific abilities (see Bae et al., 2024; Bian et al., 2018; Cheryan & Markus, 2020 for discussion) and hinder their development of science SC (Ertl et al., 2017; Lee et al., 2021). In these STEM environments, perceived effort may have different meanings for men and women, thereby informing ability differently across genders. Relevant research supports this idea. Studies on field-specific ability beliefs indicate that women tend to have more negative motivational experiences than men in fields where success is linked to brilliance rather than effort, such as STEM (Bian et al., 2018; Leslie et al., 2015; Meyer et al., 2015; Muradoglu et al., 2023). Additionally, evidence on attribution bias shows that men often attribute their successes to ability and their failures to external factors (e.g., misfortune, difficult tasks), while women are likely to attribute their success to effort or external factors (e.g., fortune, easy tasks) and their failures to a lack of ability in STEM (Kiefer & Shih, 2006; Koch et al., 2008; LaCosse et al., 2016). These findings indicate that women may be more sensitive than men to the effort they invest in STEM, as their high levels of effort might be perceived as a reflection of a lack of ability due to gender stereotypes.

Based on these findings, we hypothesized that the gendered mechanisms underlying science SC may be linked to the different dynamics of perceived effort across genders in STEM. Specifically, women's perception of putting in more effort than others may serve as critical information that confirms gendered ability stereotypes. Smith et al. (2013) provide insights into the gendered roles of social comparisons of effort. Their findings revealed that women's science SC were significantly higher when their efforts in tasks were described without comparisons ("You put in a lot of effort.") rather than in comparison to others ("You put in a lot more effort than others."). Despite its theoretical and practical significance, research exploring the gendered dynamics between perceived effort and science SC is limited. Thus, we tested this in this study.

Present Study

We aimed to identify the gendered mechanisms underlying the complex relations between perceived effort and science SC in STEM. Accordingly, our research goal was to examine whether and to what extent reciprocal relations among two types of perceived effort (criterion effort, comparative effort), science SC, and achievement (midterms, final exam) differ as a function of gender in undergraduate introductory chemistry courses. Drawing on theories (Festinger, 1954; Nicholls, 1978) and empirical evidence (Muenks et al., 2016; Smith et al., 2013), we explored perceived effort as a dimension of social comparisons. To do so, we adopted a novel approach and measured two types of perceived effort in reference to the self and others; *criterion effort* and *comparative effort*, respectively. Guided by the I/E model (Marsh & Craven, 2002) and relevant work (Müller-Kalthoff et al., 2017; Wolff et al., 2018; Wolff & Möller, 2022), we examined gender differences in the relations between these two types of perceived effort and science SC. Furthermore, we utilized the framework of REM (Marsh et al., 2018) to investigate the roles of perceived effort as both a cause and an effect of science SC and achievement. As a supplemental inquiry, we examined gender differences in reciprocal relations between science SC and achievement in chemistry, addressing the classical REM. Taken together, we employed a single-semester longitudinal panel study design to pursue three research questions.

RQ1: Whether and to what extent do reciprocal relations between two types of perceived effort and science SC differ by gender?

RQ2: Whether and to what extent do reciprocal relations between two types of perceived effort and achievement differ by gender?

RQ3: Whether and to what extent do reciprocal relations between science SC and achievement differ by gender?

Regarding comparative effort, we expected that prior comparative effort would be negatively related to subsequent science SC and achievement. Due to the lack of prior research on comparative effort as an outcome, we did not make specific hypotheses about these relations. Regarding criterion effort, based on mixed findings in the literature (Marsh et al., 2016; Xu, 2018), we hypothesized that prior criterion effort would be either positively or not significantly related to subsequent science SC and achievement, while prior science SC and achievement would be positively related to subsequent criterion effort.

Regarding the supplementary question, we expected reciprocal relations between science SC and achievement, as evidenced by abundant research on REM (e.g., Wu et al., 2021). Finally, regarding gender differences, we hypothesized that the negative relations among comparative effort, science SC, and achievement would be significantly stronger for women than men in chemistry. However, after accounting for comparative effort, the gendered roles of criterion effort in REM remain exploratory. We expect our examination will contribute to the current understanding of perceived effort and a broader picture of the gendered developmental processes underlying science SC in postsecondary STEM.

Method

Participants and Procedure

Participants were undergraduate students enrolled in three sections of an introductory chemistry course at a large public university in the Midwestern United States in the fall of 2020. This course was a prerequisite for students aiming to pursue majors in the natural sciences, certain engineering majors, and health professions programs. These courses were offered in person in prior semesters; however, most activities were delivered online during the survey administration due to the COVID-19 pandemic.

Participants completed three waves of online surveys during week 4 (Time; T1), week 10 (T2), and week 14 (T3) of the 15-week introductory chemistry course in return for homework credit. The online surveys were administered through Qualtrics survey software and took approximately 15-20 minutes each to complete. Participants were informed that participation was voluntary, with no penalties for not taking part, and they were asked to provide informed consent for the use of survey responses and university records. Participants who did not complete a survey were still invited to participate in subsequent surveys. As a result, the analytic sample consisted of 690[1] undergraduate students (61.8% female; 18.2% first-generation college students; 70% White; 74.2% first year students; M_{age} =18.7 years, SD =1.60). Among the analytic sample, 650 students (94.2%) took the T1 survey, 515 students (74.6%) took the T2 survey, and 554 students (78.8%) took the T3 survey (please see Table S1 for the summary of participants' demographic and academic characteristics, and missing data analysis). This study was approved by the university's Institutional Review Board.

Measures

Criterion effort, comparative effort, and science self-concept were assessed at three time points over the course of the semester. All items for criterion and comparative effort were modified to pertain to the chemistry course in which students were enrolled, while science SC was adapted to the domain of science. Higher scores on each variable indicate a higher endorsement of the construct. All scales measured in the current study demonstrated good

[1] A total of 725 students enrolled in the chemistry courses were invited to participate in the study, but 35 did not consent to participate. We included data from all participants who took part at least once to maximize the use of available information for estimating parameters.

internal-consistency reliabilities (see Table 1 for the reliability of the measure at each time point). The full items for the measures can be found in the supplemental materials.

Criterion Effort. The four items of the effort scale developed by Marsh et al. (2016) were utilized to evaluate students' perceptions of effort they invest in their chemistry courses in comparison to their own typical levels of effort (e.g., "In this chemistry course, I invest a lot of effort in to understanding everything."). The measure used a 5-point Likert scale ranging from 1 (*not true*) to 5 (*absolutely true*).

Comparative Effort. The four items were adapted from Smith et al. (2013), where they were referred to as "effort comparison concern". These items were developed to assess students' perceptions of effort they invest in their chemistry courses relative to effort they believe other students put in for the same course (e.g., "Compared with other students in this chemistry course, how much effort do you expend in this course?"). The measure used a 5-point Likert scale ranging from 1 (*a lot less*) to 5 (*a lot more*).

Science Self-Concept. The six items were adapted from Marsh and O'Neill (1984; Self-Description Questionnaire III) to assess students' perceived ability to succeed in science (e.g., "Science is easy for me."). The measure used a 5-point Likert scale ranging from 1 (*false*) to 5 (*true*). Two items of the science SC scale showed poor factor loadings (λ = 0.02 – 0.04) in confirmatory factor analysis with the measures from the three time points. Based on these results and the fact that these items primarily reflect the affective aspect of SC (e.g., "I'm hopeless when it comes to science", Bong & Skaalvik, 2003), we decided to exclude these two items and used the remaining four items to specify the latent construct of science SC in the current study. As a result, the science SC measure used in this study is highly reflective of the cognitive aspect of SC compared to similar constructs measured in other studies.

Achievement. Chemistry midterm and final exam scores were obtained from the course instructor[2]. Midterm 2, midterm 3, and final exam were included in the analyses as numerical scores that were calculated out of 100 points. Midterm 1 was excluded from our analyses because the exam scores were released to students before the first wave of the survey (T1) began. The final course grades were not used because they were built on all midterm scores that were used in the analyses. The final exam scores were instead used as a final performance indicator as it took place after the final survey (T3) and would reflect students' overall course performance as a cumulative exam.

Prior Achievement. Students' composite scores on the ACT exam were obtained from university records. The total scores of the SAT exam of those who did not take the ACT exam were converted into the equivalent ACT scores according to the ACT/SAT concordance tables[3]. Students who did not provide either ACT or SAT scores were classified as missing data. Missing data due to this reason was 2.9 % (N = 20).

Demographics. Demographic information was obtained through an online survey and university records. Gender and race/ethnicity were self-reported at the end of the surveys. First-generation college student status was obtained from university records, indicating that neither

---

[2] In these courses, course performance was evaluated based on three midterms, a final exam, online homework, individual and team-based recitation assessments, and lab reports. Each midterm and the final exam were graded out of 100 points. The rest of the assignments had their own grading system. Each midterm exam covered different topics. The final exam was cumulative, covering all topics that were taught through the semester. In this course, homework, recitation, and lab averages were in the 80-90% range, while exam averages were in the 55-75% range.
[3] The tables posted on the ACT website were utilized for the ACT-SAT score conversion (https://www.act.org/content/dam/act/unsecured/documents/ACT-SAT-Concordance-Information.pdf).

parent had attended college, (*0=continuing generation; 1=first generation student*). Specific race/ethnicity categories were collapsed into a binary variable (*0=majority, 1=underrepresented racial minority; URM*) indicating membership in a racial/ethnic group that is underrepresented in STEM fields (African American, Hispanic/Latino, Native American, and multi racial group) or membership in a racial/ethnic majority group in STEM (White, Asian/Asian American).

Pandemic Impact. Students' perceived negative impact of the global COVID-19 pandemic on their learning in the chemistry courses was measured and included to account for the potential confounding effect related to the pandemic. A self-developed, three items scale was administered at T1 (e.g. "Due to factors related to the COVID-19 pandemic, I face challenges that interfere with my ability to complete the work in this chemistry course."). The items used a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Analytic Plan

Preliminary analyses were conducted using SPSS 27. We utilized multiple-group latent cross-lagged panel models (CLPM) to address the primary research questions, which were conducted using M*plus* 8.6 (Muthén & Muthén, 1998–2017). CLPM estimate the *autoregressive paths* indicating the stability of a variable (e.g., the stability of criterion effort across the three time points) and the *cross-lagged paths* showing the potential causal effects of a change in a variable (e.g., criterion effort at Time 1) on change in another variable (e.g., science SC at Time 2) over the course of time. Therefore, CLPM is used to determine the directional relations between two variables (cross-lagged associations) by controlling for prior levels of the predicted variables (autoregressive associations; Kearney, 2017).

Confirmatory factor analysis (CFA) for three measures – criterion and comparative effort, and science SC – was conducted to evaluate construct validity and the full SEM model for primary research questions. We then tested for measurement invariance of three variables separately across (a) genders and (b) time points to ensure that men and women perceived the same measures in a conceptually similar way across the three time points (Putnick & Bornstein, 2016). The three measurement invariance models were tested as follows: (a) a *configural invariance* model in which the overall factor structure was restricted to be equivalent across genders and time points, respectively; (b) a *metric invariance* model in which the factor loadings were constrained to be equal across genders and time points, respectively; and (c) a *scalar invariance* model in which the equality constraints were imposed to item intercepts across genders and time points, respectively (Dimitrov, 2010). Changes in CFI greater than .010 and RMSEA greater than .015 were considered to indicate that measurement invariance was violated across genders or time points (Chen, 2007; Cheung & Rensvold, 2002).

Next, a series of multiple group latent cross-lagged panel analyses were conducted based on robust maximum likelihood estimator (MLR) in conjunction with full information maximum likelihood (FIML) to account for missing data (Enders, 2022). An array of covariates (URM, first-generation college student status, prior achievement, and pandemic impact) were included to control for the effects of potential confounding variables. Given that participants' responses were considered dependent within the same course section (McNeish et al. 2017), the nested structure of data was addressed by using type = complex on the basis of the course section in M*plus* (Muthén & Muthén, 1998–2017).

To evaluate gender differences in the path coefficients, all structural paths were first constrained to be equal between genders (fully constrained model). Then, the path was freely estimated successively by relaxing an equality constraint on a path from the fully constrained model. Chi-square difference tests were conducted to compare the fit of the freely constrained model to those of the fully constrained one (Satorra & Bentler, 2010). Significant chi-square values indicated that the path coefficients were statistically different across gender.

Results

Preliminary Results

We first examined the correlation matrix and descriptive statistics (see Table 1). We then tested the tenability of statistical assumptions for multiple-group CLMP in M*plus* 8.6. The tests confirmed the tenability of all statistical and longitudinal assumptions (see the supplemental materials for preliminary analyses results).

Primary Results

We conducted a series of multiple-group CLMP using the full analytic sample (N = 690) to address the research questions (see Figure S1 for the baseline model). Prior achievement (ACT/SAT), URM status, first-generation college student status, and the pandemic impact were included as covariates in the analyses (see Table S6). The structural model displayed an adequate fit, $\chi2$ (904) = 2017.353, p <.001, CFI = 0.925, TLI = 0.914, RMSEA =0.044 [90% CI = 0.042, 0.047], SRMR =0.085. The chi-square difference tests showed significant gender differences in autoregressive effects of comparative effort ($\Delta\chi^2$ (1) = 0.63, p < .001) and science SC ($\Delta\chi^2$ (1) =

2.38, $p < .001$) between T1 and T2[4]. Table 2 represents the parameter estimates across genders[5]; Figure 1 displays the significant parameters only.

RQ1. Reciprocal Relations Between Two Types of Perceived Effort and Science SC

The prominent role of criterion effort in science SC emerged for women, but not for men, across the three time points. Specifically, prior criterion effort was positively associated with subsequent science SC for women across the three time points ($\beta = 0.08$, $p < .05$ between T1 and T2; $\beta = 0.14$, $p < .01$ between T2 and T3), after controlling for prior science SC and an array of covariates. However, these relations were not significant for men. The negative relations between prior comparative effort and subsequent science SC were significant for both men and women at different time points. Prior comparative effort was negatively related to subsequent science SC for women between T1 and T2 ($\beta = -0.28$, $p < .01$) and for men between T2 and T3 ($\beta = -0.27$, $p < .01$), after controlling for prior science SC. In terms of the opposite direction of influence, prior science SC was not significantly associated with subsequent perceived effort for both genders across the three time points. Therefore, results partially support our hypothesis regarding gender differences, while rejecting the hypothesis about reciprocal relations between two types of perceived effort and science SC.

RQ2. Reciprocal Relations Between Two Types of Perceived Effort and Achievement

The positive and reciprocal relations between criterion effort and achievement were significant for both genders between T1 and T2. Furthermore, these relations remained significant up to T4 (final exam) for women only, thereby supporting our hypothesis.

---

[4] $\Delta\chi^2$ = Satorra-Bentler scaled chi-square value; The values less than 3.841 indicate significant gender differences in parameter estimates with one degree of freedom.
[5] Table 2 is the simple version of the results. Table S4 contains all results from the primary analyses.

Table 1 *Descriptive statistics and bivariate correlations among study variables*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Sci SC T1	1																
2. Sci SC T2	.73**	1															
3. Sci SC T3	.67**	.82**	1														
4. Cri. Eff. T1	.05	0.05	.09*	1													
5. Cri. Eff. T2	.02	.13**	.17**	.60**	1												
6. Cri. Eff. T3	.01	.11*	.16**	.55**	.72**	1											
7. Comp. Eff. T1	-.60**	-.58**	-.53**	.22**	.19**	.19**	1										
8. Comp. Eff. T2	-.54**	-.63**	-.57**	.20**	.23**	.19**	.75**	1									
9. Comp. Eff. T3	-.48**	-.57**	-.57**	.16**	.14**	.24**	.66**	.76**	1								
10. Midterm 2	.35**	.53**	.52**	0.05	.20**	.17**	-.37**	-.43**	-.39**	1							
11. Midterm 3	.30**	.41**	.49**	.08*	.20**	.26**	-.29**	-.33**	-.30**	.75**	1						
12. Final Exam	.33**	.48**	.49**	0.06	.18**	.23**	-.36**	-.39**	-.34**	.79**	.78**	1					
13. Prior Ach.	.37**	.39**	.39**	-0.07	-0.06	-0.03	-.38**	-.41**	-.36**	.50**	.42**	.51**	1				
14. Female	-.19**	-.28**	-.17**	.18**	.15**	.18**	.26**	.29**	.25**	-.19**	-.15**	-.15**	-.08*	1			
15. First-Gen.	-.17**	-.23**	-.19**	-0.07	-0.05	-0.04	.16**	.17**	.21**	-.25**	-.23**	-.25**	-.33**	0.05	1		
16. URM	-.09*	-.11**	-0.07	0.06	0.01	-0.00	.12**	.16**	.16**	-.13**	-.10**	-.13**	-.21**	0.05	.16**	1	
17. Pandemic	-.14**	-.17**	-.13**	-.11**	-0.08	-0.02	.19**	.17**	.19**	-.19**	-.16**	-.14**	-.10**	0.04	.10**	-0.02	1
n	650	514	544	650	512	542	650	512	542	675	673	673	670	690	690	690	690
M	3.54	3.44	3.45	4.33	4.19	4.14	3.31	3.33	3.28	74.32	74.9	71.5	28.54	.62	.18	.16	3.22
SD	.70	.75	.75	.57	.70	.76	.85	.83	.82	18.14	18.67	17.53	3.76	.48	.38	.36	1.11
Kurtosis	.40	.44	.94	.48	.75	.30	-.18	-.15	.05	-.23	.60	.71	.91	-1.75	.71	1.42	-.82
Skewness	-.32	-.44	-.55	-.71	-.84	-.83	-.11	.04	.06	-.68	-.93	-.87	-.87	-.49	1.64	1.85	-.12
Min.	1	1	1	1	1	1	1	1	1	12.5	0	0	15	0	0	0	1
Max.	5	5	5	5	5	5	5	5	5	100	100	100	36	1	1	1	5
α	.89	.90	.90	.85	.90	.91	.85	.84	.86	-	-	-	-	-	-	-	.89

Notes. Sci SC = Science self-concept; Cri. Eff. = Criterion effort; Comp. Eff .= Comparative effort; Prior Ach = Prior achievement (ACT/SAT); First-Gen = First-generation college student status; URM = Underrepresented minority student status; Pandemic = Pandemic impact. * $p < .05$. ** $p < .01$. *** $p < .001$.

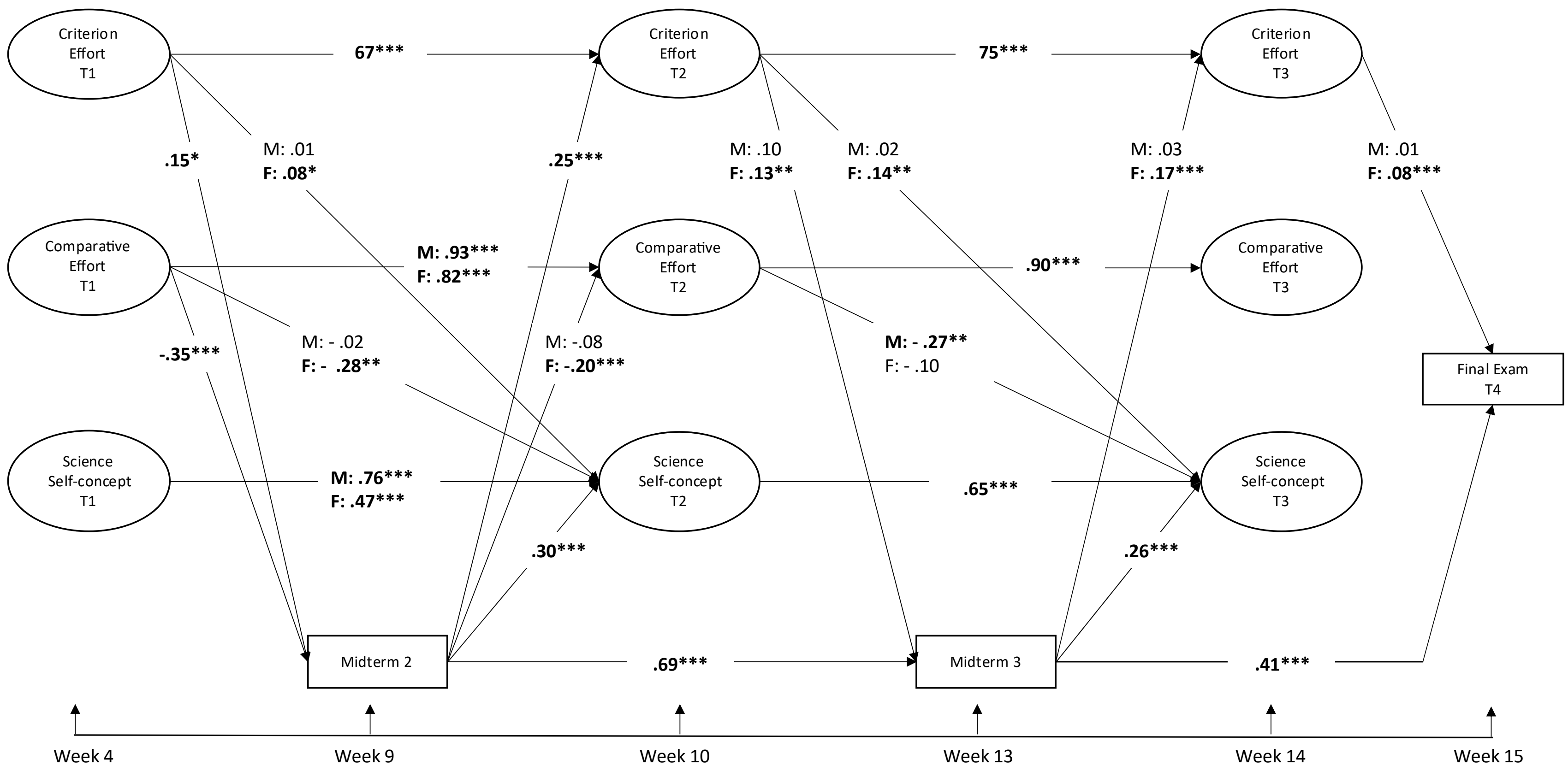
Table 2
Summary of path coefficients among study variables across genders

Predictors	Outcomes	T1 → T2				T2 → T3			
		Male		Female		Male		Female	
		β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Criterion effort	Criterion effort	**.63*** **	**.09**	**.67*** **	**.02**	**.79*** **	**.14**	**.73*** **	**.10**
	Comp. effort	-.02	.13	.03	.04	-.08	.11	-.04	.08
	Science SC	.01	.03	**.08** **	**.02**	.02	.03	**.14** **	.08
	Midterm 2	**.15* **	**.08**	**.17* **	**.07**	-	-	-	-
	Midterm 3	-	-	-	-	.10	.08	**.13** **	**.02**
Comp. effort	Criterion effort	.09	.16	.08	.06	-.00	.15	.12	.20
	Comp. effort	**.93*** **	**.14**	**.82*** **	**.06**	**.91*** **	**.06**	**.87*** **	**.06**
	Science SC	-.02	.18	**-.28*** **	**.13**	**-.27** **	**.11**	-.10	.15
	Midterm 2	**-.25** **	**.09**	**-.36*** **	**.12**	-	-	-	-
Science SC	Criterion effort	.14	.18	.06	.11	-.09	.13	.06	.16
	Comp. effort	-.09	.13	-.00	.05	-.03	.03	-.01	.11
	Science SC	**.76*** **	**.14**	**.47*** **	**.06**	**.60*** **	**.17**	**.67*** **	**.17**
Midterm 2	Criterion effort	**.22*** **	**.10**	**.25*** **	**.02**	-	-	-	-
	Comp. effort	-.08	.09	**-.20*** **	**.05**	-	-	-	-
	Science SC	**.28** **	**.05**	**.35*** **	**.05**	-	-	-	-
Midterm 3	Criterion effort	-	-	-	-	.03	.11	**.17*** **	**.02**
	Comp. effort	-	-	-	-	-.06	.03	-.04	.06
	Science SC	-	-	-	-	**.27*** **	**.04**	**.26** **	**.09**

Notes. Significant coefficients are in bold. The full display of path coefficients is shown in Table S4. Missing estimates indicate that the paths were not tested due to the temporal order issue. Midterm 2 took place between T1 and T2 and midterm 3 occurred between T2 and T3. Midterm 1 was excluded from the analyses (see the Measures section). Comp. effort = Comparative effort. *$p < .05$. ** $p < .01$. *** $p < .001$.

Figure 1

Standardized path coefficients of the study model.



Notes. Only significant paths from the model are shown in the figure. The path from midterm 2 to final exam was significant but omitted in the figure for simplicity. Paths with multiple coefficients indicate that gender differences existed (M = Male, F = Female). Paths with a single coefficient indicate that gender differences were not significant thus the path coefficients from a single group analysis are shown for simplicity. Correlated errors among study variables and covariates are not shown for simplicity. The full display of path coefficients is shown in Table S4. * $p < .05$. ** $p < .01$. *** $p < .001$.

regarding criterion effort. Specifically, criterion effort at T1 was positively related to midterm 2 (Male: $\beta = 0.15$, $p < .05$; Female: $\beta = 0.17$, $p < .05$), after controlling for prior achievement. Midterm 2 was also positively related to criterion effort at T2 (Male: $\beta = 0.22$, $p < .001$; Female: $\beta = 0.25$, $p < .001$), after controlling for criterion effort at T1. For women only, criterion effort at T2 was positively related to midterm 3 ($\beta = 0.13$, $p < .01$), and midterm 3 was positively related to criterion effort at T3 ($\beta = 0.17$, $p < .001$). Lastly, women's criterion effort at T3 was positively related to final exam at T4 ($\beta = 0.08$, $p < .001$).

The negative and reciprocal relations between comparative effort and achievement were significant for women only between T1 and T2. Although comparative effort at T1 was negatively related to midterm 2 for both genders (Male: $\beta = - 0.25$, $p < .01$; Female: $\beta = - 0.36$, $p < .001$), midterm 2 was negatively linked to comparative effort at T2 ($\beta = - 0.20$, $p < .001$) for women only. The relations between comparative effort and achievement were not significant for both genders between T2 and T3. Overall, our hypothesis regarding this research question was supported.

RQ3. Reciprocal Relations Between Science SC and Achievement

Chemistry achievement emerged as a significant precursor of science SC for both genders across the three time points; however, the relation in the opposite direction was not significant. For both genders, midterm 2 was positively related to science SC at T2 (Male: $\beta = 0.28$, $p < .01$; Female: $\beta = 0.35$, $p < .001$), and midterm 3 was positively linked to science SC at T3 (Male: $\beta = 0.27$, $p < .001$; Female: $\beta = 0.26$, $p < .01$). Contrary to our hypothesis, science SC was not predictive of achievement for either gender. Thus, the results did not support our hypothesis concerning this research question.

Discussion

In this study, we examined whether and to what extent reciprocal relations among two types of perceived effort (criterion, comparative effort), science SC, and achievement (midterms, final exam) differ as a function of gender in undergraduate introductory chemistry. To address this inquiry, we uniquely integrated and extended Nicholls' work (1978) and two main theories in the ASC literature (I/E model; Marsh & Craven, 2002; REM; Marsh et al., 2018) by examining two types of perceived effort in the framework of REM. The findings from the current study will (a) enhance the theoretical understanding of the relations between perceived effort and science SC and (b) provide insights on how to support women in leveraging their efforts to achieve their goals in stereotyped fields during their college transitions.

Differential Roles of Comparative Effort and Criterion Effort in Science Self-Concept

It is noteworthy that the predictive validity of comparative effort for science SC was significant above and beyond that of criterion effort among college students in introductory chemistry. Our findings corroborate and extend prior work on the negative relations between comparative effort (i.e., the perception of trying harder than others) and ASC (Jagacinski & Nicholls, 1987; Muenks et al., 2016; Nicholls, 1978) by testing the utility of comparative effort for science SC. Building on research on social comparison theory (Boissicat et al., 2021) and the relations between effort and ability (Jagacinski & Nicholls, 1987; Muenks et al., 2016), our findings suggest that students tend to consider relative effort when assessing their scientific abilities. Moreover, both genders tend to interpret relatively higher effort than others as a sign of low scientific ability (e.g., "I am not good at science because I should work harder than others in science class.").

Drawing upon prior findings (Inzlicht et al., 2018; Muenks et al., 2016), social comparisons of effort may relate to the perception of effort as a response to subjective task difficulty. When students recognize that their effort is greater than that of their peers, they may view this additional effort as needed to compensate for a lack of ability to tackle challenging tasks. Consistent with the argument by Jagacinski and Nicholls (1987), these negative associations may be influenced by the social context, which can make them particularly pronounced in the current study – i.e., the competitive introductory chemistry course. Thus, future research should further investigate these associations across different fields or cultures, particularly in contexts that are less competitive or that place a higher value on effort (Leslie et al., 2015; O'Donnell et al., 2023; Xu, 2018).

In terms of gender differences, these negative associations manifested at different time points - between T1 and T2 for women, and between T2 and T3 for men. Supported by research on attribution bias (LaCosse et al., 2016), our findings suggest that women may more promptly focus on their lack of ability to explain relatively higher effort they expend, while men might seek external reasons to understand their comparative effort early in the semester. However, our results indicate that the significance of comparative effort may become pronounced for men as the workload increases over the semester, leading them to question whether their abilities are not as superior as presumed or stereotyped in STEM. Overall, our findings indicate that the correlates of comparative effort may be significant for both genders in STEM, as gendered stereotypes about ability are either affirmed (for women) or challenged (for men) by the greater effort they put in compared to their peers in STEM. Future research should explore this potential explanation by examining these relations across STEM fields that are associated with varying levels of gender stereotypes (Cheryan et al., 2017) and brilliance orientation (Bian et al., 2018).

Notably, the gendered role of criterion effort in women's science SC was evident. We found that criterion effort (i.e., the perception of trying hard) was positively related to women's science SC after accounting for comparative effort. However, these relations were not significant for men. This gendered prediction is supported by evidence showing that women tend to view effort as the means to success in STEM (success/effort attribution; LaCosse et al., 2016). Moreover, prior research indicates that women may perceive effort they invest as a reflection of their dedication to achieving educational and career goals in STEM (cf. Inzlicht et al., 2018; Oyserman et al., 2018; Smith & Oyserman, 2015). Specifically, undergraduate women in the current study might have recognized that pursuing STEM majors requires considerable effort to overcome gender stereotypes. As a result, they may not view high levels of effort (criterion effort) as an indication of a lack of ability. Instead, they may see their criterion effort as reflective of the values or importance they place on this field (cf., Inzlicht et al., 2018; Jelsma et al., 2022; Oyserman et al., 2018), which may positively relate to science SC.

Conversely, criterion effort was not directly related to men's science SC. Given that men generally exhibit higher competence than women in STEM (Jansen et al., 2014, 2015), men's science SC may be more durable and thus more likely to be influenced by objective measures of ability (e.g., achievement) rather than subjective indicators (e.g., perceived effort) in STEM. Although it was not the primary research question of the current study, our results revealed that two types of perceived effort were significantly associated with men's science SC through chemistry achievement. These findings indicate that achievement feedback may be particularly critical for men to understand the relations between effort and ability in STEM, aligning with Covington's (1984) claim. Thus, further research exploring the potential role of achievement in the gendered relations between perceived effort and science SC is warranted.

Unexpectedly, science SC was not predictive of two types of perceived effort for either gender across the three time points. These findings seem counterintuitive considering the theoretical arguments and empirical evidence on positive relations between ASC and effort (Covington, 1984). However, some previous research has reported similar results in math (Xu, 2018). Specifically, it is possible that students with high competence may wish to align themselves with the prototype of a 'brilliant person' in STEM (Bian et al., 2018; Cheryan et al., 2013; Hannover & Kessels, 2004), which may lead them to be reluctant to acknowledge that they try hard or harder than their peers. In this context, students' science SC may not necessarily relate to their subsequent perceived effort. Another plausible explanation is that objective achievement would more effectively inform perceived effort. For example, favorable prior grades may better lead individuals to exert effort to maintain or improve their future performance. Indeed, findings from the present study support this perspective, as discussed in the following section. Taken together, our findings provide a more nuanced understanding of the relations between perceived effort and science SC by illustrating the contrasting roles of criterion effort and comparative effort, particularly in the stereotyped fields.

Reciprocal Relations Between Perceived Effort and Achievement

We found that two types of perceived effort and achievement were related reciprocally, with more pronounced patterns among women. Positive reciprocity between criterion effort and achievement were found for both genders between T1 and T2, and these relations were presented until T4 for women. Specifically, students' criterion effort (i.e., the perception of trying hard) were linked to high midterm scores, which in turn were related to an increase in their criterion effort. Consistent with prior research (Xu, 2018; cf., Inzlicht et al., 2018), our findings indicate that subsequent effort can be reinforced when individuals experience success resulting from their prior high effort. In contrast, while prior comparative effort was negatively related to chemistry performance for both genders, negative reciprocity emerged among women only between T1 and T2. Our findings suggest that women may experience negative feedback loops between comparative effort and achievement, where their perception of putting in more effort than others – potentially linked to subjective task difficulty - is intensified by receiving low grades in STEM. These findings highlight the importance of early successful experiences for undergraduate students, particularly women, in their understanding of the effort they invest within the context of introductory chemistry courses.

Protective Role of Achievement in Science Self-Concept

Chemistry midterm scores positively predicted science SC for both genders across the three time points. Unexpectedly, however, the opposite directional relations were not significant. Although these results seem to contradict prior evidence on REM, there are several possible explanations. Science SC may heavily depend on achievement during academic transitions as the frame of reference changes (Valentine et al., 2004). Another possibility is that, given the current study's design, the explanatory power of perceived effort outweighs that of science SC for STEM achievement in general or among undergraduates in this study in particular. To our best knowledge, this is the first study to explore the relations between perceived effort, science SC, and achievement in postsecondary STEM settings. Additional research is needed to further investigate the initial trends noted in the present study.

Implications, Limitations, and Recommendations for Future Research

The current study provides several implications for gender equity in STEM. First, our findings suggest that parents and educators should encourage a self-referenced view and discourage an other-referenced view on effort. This approach can be an effective way to support science SC and success of *both* genders in introductory chemistry. Prior work (Nicholls, 1984) indicates that comparative effort is associated with a performance orientation focused on external standards, while criterion effort aligns with a mastery orientation centered on internal or self-set standards. In this sense, students' criterion effort can be enhanced by emphasizing the importance of mastering chemistry course content.

Second, our results suggest that providing students, particularly women, with diverse opportunities for success are important in introductory chemistry, as early successful experiences can help students create positive feedback loops between criterion effort and performance, as well as enhance their science SCs. Third, our findings contribute to the theoretical understanding of the complex relations between perceived effort and science SC by providing empirical evidence on comparative effort. This study highlights the unique roles of comparative effort in science SC above and beyond that of criterion effort in introductory chemistry. Further, gendered roles of two perceived efforts in science SC and achievement extend our understanding of gendered motivational pathways in undergraduate STEM.

Despite the aforementioned implications, this study has several limitations; thus, suggestions are made for future research. First, we focused on students' science SC, rather than chemistry

SC, and we investigated its relations with perceived effort and achievement in a specific chemistry course. This decision was deliberate, as most participants were first year students who may not have developed ASC in a particular STEM domain before declaring their majors. Thus, future research is encouraged to examine these constructs within the same domain to provide a deeper understanding of the relations between two types of perceived effort and ASC. Second, general others, rather than a particular person, were established as the comparison standard in this study. Although this study design is supported by prior studies (e.g., Zell & Alicke, 2020), the assimilation effect (Huguet et al., 2009) points to the possibility that students may have selected and compared themselves with a subset of peers who are similar to them. Therefore, future research should further examine the processes of effort comparisons in relation to specific comparison targets to expand upon findings from the current study. Third, it should be noted that the current study was conducted during the COVID-19 pandemic (Fall 2020). Although we accounted for the challenges students perceived during this time through statistical controls (please see the supplemental materials for the full items and Table S6 for the coefficients), our approach may not have fully detected the influences of the COVID-19 pandemic that were associated with changes in course settings and broader social contexts (e.g., remote learning, social distancing; Fong et al., 2022). Thus, our examination during this time may have narrowly captured the associations among study variables. Future research should be conducted to replicate these relations and complement the findings of the current study.

Fourth, building on the current study's evidence, future research should further examine the contextual/individual factors that influence the salience and dynamics of two types of perceived effort, and extend this examination into academic self-concept in different domains. Fifth, the specific nature of the sample in this study limits the generalizability of our findings to undergraduate students in introductory chemistry courses. It is possible that social comparison of effort may have been particularly pronounced in the introductory STEM course settings, characterized by college transitions and competitive course climates. Therefore, there is a critical need to pursue this line of investigation in cross-cultural settings (e.g., comparing Western and Asian cultures in terms of their differing cultural values on effort, O'Donnell et al., 2023), in different fields (e.g., comparing fields that value effort to varying degrees), or among diverse student populations (e.g., comparing students in non- or different STEM fields) to improve the generalizability of these findings. Next, while our approach to the study sample and sampling maximizes statistical power and supports our focus on the transitional and competitive nature of the introductory course context, it is important to note that nearly 25% of the sample consisted of non-first-year students. These students may have experienced different motivational processes, as their course commitments and the perceptions associated with those commitments could differ from those of first-year students. Future research should extend the current study by focusing heavily on first-year students or by exploring different associations among students from varying academic ranks or educational stages. Finally, all constructs in the current study, except for achievement, were measured using retrospective self-report measures. Although all measures were previously validated and demonstrated excellent reliability, future research should consider measuring perceived effort and science SC more holistically by adopting qualitative methods, mixed methods design, or incorporating other data sources (e.g., direct observation, teacher-, peer-, and parent reports).

Conclusion

This study examined whether and to what extent reciprocal relations between two types of perceived effort (criterion, comparative effort), science SC, and achievement differ as a function of gender in introductory chemistry. Key findings included: (a) distinct roles of two types of perceived effort in science SC, (b) reciprocal relations between these perceptions of effort and achievement, and (c) gender differences in these associations. Our results suggest that promoting criterion effort (self-referenced views on effort) and downplaying comparative effort (other-referenced views on effort) can enhance both genders' science SC and success in STEM fields.

References

Bae, C. L., Matewos, A. M., & Fife, J. (2024). Supporting student voice in science classrooms: The limits of psychosocial approaches and the importance of sociocultural and critical perspectives on student agency. *Educational Psychologist*, 1-21. https://doi.org/10.1080/00461520.2024.2370549

Bian, L., Leslie, S. J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist*, *73*(9), 1139–1153. https://doi.org/10.1037/amp0000427

Boissicat, N., Fayant, M. P., Nurra, C., & Muller, D. (2022). Social comparison in the classroom: Priming approach/avoidance changes the impact of social comparison on self-evaluation and performance. *British Journal of Educational Psychology, 92*(2), e12466. DOI: 10.1111/bjep.12466

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really?. *Educational Psychology Review, 15*, 1-40. https://doi.org/10.1023/A:1021302408382

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*. Retrieved from: chromeextension://efaidnbmnnnibpcajpcglclefindmkaj/https://files.eric.ed.gov/fulltext/ED544470.pdf

Cheryan, S., & Markus, H. R. (2020). Masculine defaults: Identifying and mitigating hidden cultural biases. *Psychological Review, 127*(6), 1022–1052. https://doi.org/10.1037/rev0000209

Cheryan, S., Plaut, V. C., Handron, C., & Hudson, L. (2013). The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex Roles: A Journal of Research, 69*(1-2), 58–71. https://doi.org/10.1007/s11199-013-0296-x

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1–35. https://doi.org/10.1037/bul0000052

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Chiu, M.-S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology, 104*(1), 87–107. https://doi.org/10.1037/a0025734

Covington, M. V. (1984). The self-worth theory of achievement motivation: Findings and implications. *The Elementary School Journal, 85*(1), 5–20. https://doi.org/10.1086/461388

Denice, P. A. (2021). Choosing and changing course: Postsecondary students and the process of selecting a major field of study. *Sociological Perspectives, 64*(1), 82-108. https://doi.org/10.1177/0731121420921903doi:0731121420921903

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research, 78*(4), 828–879. https://doi.org/10.3102/0034654308321210

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development.* Philadelphia: Psychology Press.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Enders, C. K. (2022). *Applied missing data analysis*. New York, NY: Guilford Press.

Ertl, B., Luttenberger, S., & Paechter, M. (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Frontiers in Psychology*, *8*, 703. doi: 10.3389/fpsyg.2017.00703

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Fong, C. J. (2022). Academic motivation in a pandemic context: A conceptual review of prominent theories and an integrative model. *Educational Psychology*, *42*(10), 1204-1222. https://doi.org/10.1080/01443410.2022.2026891

Gibbons, R. E., & Raker, J. R. (2019). Self-beliefs in organic chemistry: Evaluation of a reciprocal causation, cross-lagged model. *Journal of Research in Science Teaching, 56*(5), 598-618. https://doi.org/10.1002/tea.21515

Gniewosz, B., Eccles, J. S., & Noack, P. (2012). Secondary school transition and the use of different sources of information for the construction of the academic self-concept. *Social Development, 21*(3), 537–557. https://doi.org/10.1111/j.1467-9507.2011.00635.x

Hannover, B., & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices. Why high school students do not like math and science. *Learning and Instruction, 14*(1), 51–67. https://doi.org/10.1016/j.learninstruc.2003.10.002

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., ... & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*(1), 156. DOI: 10.1037/a0015558

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337-349. DOI: 10.1016/j.tics.2018.01.007

Jagacinski, C. M., & Nicholls, J. G. (1987). Competence and affect in task involvement and ego involvement: The impact of social comparison information. *Journal of Educational Psychology, 79*(2), 107–114. https://doi.org/10.1037/0022-0663.79.2.107

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

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

Jelsma, E., Chen, S., & Varner, F. (2022). Working harder than others to prove yourself: High-effort coping as a buffer between teacher-perpetrated racial discrimination and mental health among Black American adolescents. *Journal of Youth and Adolescence, 51*(4), 694–707. https://doi.org/10.1007/s10964-021-01563-4

Kearney, M. W. (2017). Cross-lagged panel analysis. In M. R. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods* (pp. 312–314). Thousand Oaks, CA: SAGE. https://doi.org/10.4135/9781483381411.n117

Kiefer, A., & Shih, M. (2006). Gender differences in persistence and attributions in stereotype relevant contexts. *Sex Roles: A Journal of Research, 54*(11-12), 859–868. https://doi.org/10.1007/s11199-006-9051-x

Kim, Y. E., Yu, S. L., Koenka, A. C., Lee, H., & Heckler, A. F. (2022). Can self-efficacy and task values buffer perceived costs? Exploring introductory-and upper-level physics courses. *The Journal of Experimental Education*, *90*(4), 839-861. https://doi.org/10.1080/00220973.2021.1878992

Koch, S. C., Müller, S. M., & Sieverding, M. (2008). Women and computers. Effects of stereotype threat on attribution of failure. *Computers & Education, 51*(4), 1795-1803. https://doi.org/10.1016/j.compedu.2008.05.007

LaCosse, J., Sekaquaptewa, D., & Bennett, J. (2016). STEM stereotypic attribution bias among women in an unwelcoming science setting. *Psychology of Women Quarterly, 40*(3), 378-397. https://doi.org/10.1177/0361684316630965

Lee, H., Du, W., Pedersen, R. M., Estrada, M., Adams, A. S., Barnes, R. T., ... & Hernandez, P. R. (2024). To stay, switch, or leave: A four-year longitudinal study of the situated and stable social influences on women's STEM major choices. *Contemporary Educational Psychology*, *79*, 102324. https://doi.org/10.1016/j.cedpsych.2024.102324

Lee, H., Shirley, L. Y., Kim, M., & Koenka, A. C. (2021). Concern or comfort with social comparisons matter in undergraduate physics courses: Joint consideration of situated expectancy-value theory, mindsets, and gender. *Contemporary Educational Psychology, 67*, 102023. https://doi.org/10.1016/j.cedpsych.2021.102023

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262-265. DOI: 10.1126/science.1261375

Liou, P. Y., Wang, C. L., Lin, J. J., & Areepattamannil, S. (2021). Assessing students' motivational beliefs about learning science across grade level and gender. *The Journal of Experimental Education*, *89*(4), 605–624. https:// doi.org/10.1080/00220973.2020.1721413

Marsh, H. W. (1993). Academic self-concept: Theory, measurement, and research. In J. M. Suls (Vol. Ed.), The self in social perspective. Vol. 4. *Psychological perspectives on the self* (pp. 59–98). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Marsh, H. W., & Craven, R. (2002). The pivotal role of frames of reference in ASC formation: The big fish little pond effect. In F. Pajares, & T. Urdan (Eds.) *Adolescence and education* (vol. II, pp. 83–123). Greenwich, CT: Information Age.

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21*(2), 153–174. https://doi.org/10.1111/j.1745-3984.1984.tb00227.x

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

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology, 54*(2), 263–280. https://doi.org/10.1037/dev0000393

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140. https://doi.org/10.1037/met0000078

Meaders, C. L., Lane, A. K., Morozov, A. I., Shuman, J. K., Toth, E. S., Stains, M., ... & Smith, M. K. (2020). Undergraduate student concerns in introductory STEM courses: What they are, how they change, and what influences them. *Journal for STEM Education Research*, *3*(2), 195-216. https://doi.org/10.1007/s41979-020-00031-1

Meyer, M., Cimpian, A., & Leslie, S. J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology, 6*, 235. https://doi.org/10.3389/fpsyg.2015.00235

Muenks, K., & Miele, D. B. (2017). Students' thinking about effort and ability: The role of developmental, contextual, and individual difference factors. *Review of Educational Research*, *87*(4), 707-735. https://doi.org/10.3102/0034654316689328

Muenks, K., Miele, D. B., & Wigfield, A. (2016). How students' perceptions of the source of effort influence their ability evaluations of other students. *Journal of Educational Psychology, 108*(3), 438–454. https://doi.org/10.1037/edu0000068

Müller-Kalthoff, H., Helm, F., & Möller, J. (2017). The big three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Social Psychology of Education: An International Journal, 20*(4), 849–873. https://doi.org/10.1007/s11218-017-9395-9

Muradoglu, M., Arnold, S. H., Leslie, S. J., & Cimpian, A. (2023). "What does it take to succeed here?": The belief that success requires brilliance is an obstacle to diversity. *Current Directions in Psychological Science, 32*(5), 379-386. https://doi.org/10.1177/09637214231173361

Muthén, L. K. & Muthén, B. O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: The big-fish-little-pond-effect for

PISA 2006. *Educational Psychology*, *31*(5), 629-656. https://doi.org/10.1080/01443410.2011.586416

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104*(4), 1033–1053. https://doi.org/10.1037/a0027697

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and Course Selection in Upper Secondary Education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12*(4), 323–345. https://doi.org/10.1080/13803610600765687

National Center for Education Statistics (2021). *Integrated Postsecondary Education Data System (IPEDS)*, Fall 2011 through Fall 2020, Completions component. Available at https://nces.ed.gov/programs/digest/d21/tables/dt21_318.45.asp

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49*(3), 800–814. https://doi.org/10.2307/1128250

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*(3), 328–346. https://doi.org/10.1037/0033-295X.91.3.328

O'Donnell, S. C., Yan, V. X., Bi, C., & Oyserman, D. (2023). Is difficulty mostly about impossibility? What difficulty implies may be culturally variant. *Personality and Social*

Psychology Bulletin, *49*(2), 309-328. https://doi.org/10.1177/01461672211065595

Oyserman, D., Elmore, K., Novin, S., Fisher, O., & Smith, G. C. (2018). Guiding people to interpret their experienced difficulty as importance highlights their academic possibilities and improves their academic performance. *Frontiers in Psychology*, *9*, 781. https://doi.org/10.3389/fpsyg.2018.00781

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology, 84*(1), 152–174. https://doi.org/10.1111/bjep.12028

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Robnett, R. D. (2016). Gender bias in STEM fields: Variation in prevalence and links to STEM self-concept. *Psychology of Women Quarterly, 40*(1), 65–79. https://doi.org/10.1177/0361684315596162

Saß, S., & Kampa, N. (2019). Self-concept profiles in lower secondary level–an explanation for gender differences in science course selection?. *Frontiers in Psychology*, *10*, 836. https://doi.org/10.3389/fpsyg.2019.00836

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Seymour E., & Hunter A. B. (2019). *Talking about leaving revisited: Persistence, relocation, and loss in undergraduate STEM education*. Springer. Nature: Switzerland AG.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*(1), 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Skaalvik, E. M., & Skaalvik, S. (2013). School goal structure: Associations with students' perceptions of their teachers as emotionally supportive, academic self-concept, intrinsic motivation, effort, and help seeking behavior. *International Journal of Educational Research*, *61*, 5-14. https://doi.org/10.1016/j.ijer.2013.03.007

Smith, G. C., & Oyserman, D. (2015). Just not worth my time? Experienced difficulty and time investment. *Social Cognition, 33*(2), 85–103. https://doi.org/10.1521/soco.2015.33.2.1

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin, 39*(2), 131-143. https://doi.org/10.1177/0146167212468332

Taskinen, P. H., Schütte, K., & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: The role of school factors, individual interest, and science self-concept. *Educational Research and Evaluation*, *19*(8), 717-733. https://doi.org/10.1080/13803611.2013.853620

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relations between self-beliefs and academic achievement: A systematic review. *Educational Psychologist, 39*, 111–133. doi:10.1207/s15326985ep3902_3

Villafane, S. M., Xu, X., & Raker, J. R. (2016). Self-efficacy and academic performance in first-semester organic chemistry: Testing a model of reciprocal causation. *Chemistry Education Research and Practice, 17*(4), 973-984. DOI: 10.1039/C6RP00119J

von Keyserlingk, L., Becker, M., & Jansen, M. (2019). Academic self-concept during the transition to upper secondary school. *Contemporary Educational Psychology, 56,* 152–160. https://doi.org/10.1016/j.cedpsych.2019.01.001

Wang, J., Oliver, J. S., & Staver, J. R. (2008). Self-concept and science achievement: Investigating a reciprocal relation model across the gender classification in a crosscultural context. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *45*(6), 711-725. https://doi.org/10.1002/tea.20182

Wilkins, J. L. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education*, *72*(4), 331-346. https://doi.org/10.3200/JEXE.72.4.331-346

Wolff, F., & Möller, J. (2022). An individual participant data meta-analysis of the joint effects of social, dimensional, and temporal comparisons on students' academic self-concepts. *Educational Psychology Review, 34*(4), 2569–2608. https://doi.org/10.1007/s10648-022-09686-1

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology, 110*(7), 1005–1025. https://doi.org/10.1037/edu0000248

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review,* 1-30. https://doi.org/10.1007/s10648-021-09600-1

Xu, J. (2018). Reciprocal effects of homework self-concept, interest, effort, and math achievement. *Contemporary Educational Psychology, 55,* 42–52. https://doi.org/10.1016/j.cedpsych.2018.09.002

Xu, J. (2020). Longitudinal effects of homework expectancy, value, effort, and achievement: An empirical investigation. *International Journal of Educational Research, 99*, 101507. https://doi.org/10.1016/j.ijer.2019.101507

Zell, E., & Alicke, M. D. (2020). Effects of local and general comparisons on self-assessment. In J. Suls, R. L. Collins, & L. Wheeler (Eds.), *Social comparison, judgment, and behavior* (pp. 141–175). Oxford University Press.

Zhang, F., Bae, C. L., & Broda, M. (2022). Science self-concept, relatedness, and teaching quality: A multilevel approach to examining factors that predict science achievement. *International Journal of Science and Mathematics Education*, *20*(3), 503-529. https://doi.org/10.1007/s10763-021-10165-2

Zhao, M., Ozturk, E., Law, F., Joy, A., Deutsch, A. R., Marlow, C. S., … & Rutland, A. (2024). Reciprocal associations between science efficacy, STEM identity and scientist career interest among adolescent girls within the context of informal science learning. *Journal of Youth and Adolescence, 53*(2), 472–484. https://doi.org/10.1007/s10964-023-01868-6