Considering Statistically Equivalent Models when using Structural Equation Modeling: an Example from Physics Identity

Yangqiuting Li[a] and Chandralekha Singh[a]

Corresponding author: Yangqiuting Li (yal133@pitt.edu)
[a]Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, United States

Keywords: equity, gender, identity, self-efficacy, perceived recognition

Abstract

Structural equation modeling (SEM) is a statistical method widely used in educational research to investigate relationships between variables. Using a SEM model involves a crucial step of considering statistically equivalent models and contemplating why the proposed model should not be rejected in favor of equivalent ones. However, many studies using SEM did not explicitly discuss this step. In this study, we use physics identity model as an example to demonstrate how multiple statistically equivalent models have distinct instructional implications. Previous research has indicated that physics identity comprises three dimensions: perceived recognition, self-efficacy, and interest. However, the relationships between these dimensions have not been thoroughly understood. Here, we discuss how our proposed model with perceived recognition predicting self-efficacy and interest is supported by prior studies involving individual student interviews and how intervention studies can further determine a more accurate causal model. Our study highlights the importance of considering statistically equivalent models when using SEM as an analysis tool.

Introduction and theoretical framework

Structural Equation Modeling (SEM) is a statistical method widely used in research to analyze predictive relationships among variables (Kline, 2015). It enables researchers to examine both the measurement properties of latent variables and the structural relationships between them (Kline, 2015). SEM has proven to be highly valuable across diverse fields (Kline, 2015). In physics education research, for example, SEM helps researchers to explore the predictive relationships between various factors and students' learning outcomes (Godwin, Potvin, Hazari, & Lock, 2016; Li & Singh, 2021). The process of conducting SEM analyses typically involves several steps: model specification, evaluation of model identification, data collection, and model estimation and evaluation (Kline, 2015). Within the model estimation and evaluation step, one crucial element is considering statistically equivalent models and contemplating why the proposed model should not be rejected in favor of these alternatives (Kline, 2015). However, this element is often overlooked in SEM studies, potentially undermining the robustness of research findings (MacCallum & Austin, 2000).

In this study, we use physics identity model as an example to discuss multiple statistically equivalent SEM models to the one we selected based upon other data from individual interviews. Each equivalent model may have distinct instructional implications. The remainder of the introduction is structured as follows: Firstly, we provide an overview of SEM. Then, we discuss the application of SEM to investigate the relationships among the three dimensions of physics identity: perceived recognition, self-efficacy, and interest, highlighting the significance of considering statistically equivalent models. Finally, we propose a model based on prior studies involving interviews conducted by us and findings by other researchers.

General approach to SEM

SEM involves several key steps to analyze predictive relationships among variables. The first step is model specification, in which researchers establish the hypothesized relationships between observed variables and their corresponding latent variables, as well as the relationships among the latent variables (Kline, 2015). This step relies on previous research and domain knowledge (Kline, 2015), providing the foundation for the subsequent steps of the analysis.

After specifying the model, researchers evaluate its identification (Kline, 2015). Model identification refers to determining whether the model is under-identified, just-identified, or over-identified (Kline, 2015). An under-identified model is characterized by having more parameters to be estimated than the number of data points (variances and covariances of the observed variables) available in the model. This results in negative degrees of freedom, indicating that there is insufficient information for evaluating the model. The degrees of freedom are obtained by subtracting the number of parameters to be estimated from the total number of data points. A just-identified model has an equal number of parameters to be estimated and data points. An over-identified model, on the other hand, has more data points than parameters to be estimated. A model must be just-identified or overidentified to estimate parameters (Kline, 2015). Over-identified models are particularly important in SEM because they enable researchers to examine indices of model fit, which assess how well the tested model describes the observed data (Kaplan, 2008).

Subsequently, researchers proceed to data collection, which involves gathering, preparing, and screening the data to ensure its quality and suitability for analysis (Kline, 2015). Appropriate data selection and preparation are essential for obtaining reliable and valid results.

Once the data is prepared, researchers estimate and evaluate the model using statistical software. The first step in estimation is evaluating the fit of the model to the data (Kline, 2015). Fit indices such as the chi-square test, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA), are used to assess how well the model aligns with the observed data. If the fit is unsatisfactory, researchers may need to revise the model to improve its alignment (Kline, 2015). Assuming an acceptable fit, researchers proceed to interpret the parameter estimates, which represent the strength and direction of relationships between variables (Kaplan, 2008). In addition to interpreting the proposed model, one crucial element of model evaluation is considering statistically equivalent models that provide alternative representations of the data (Kline, 2015). Statistically equivalent models have different causal structure but are identical in fit to the data (Kline, 2015). They are a set of models that yield identical correlation matrices, fit function and chi-square value, and goodness-of-fit indices (Hershberger & Marcoulides, 2006). Consequently, these equivalent models can explain the data equally well compared to researchers' preferred model but may make different causal claims (Kline, 2015). Therefore, researchers should explicitly acknowledge the existence of equivalent models, generate plausible alternative versions of their model, and provide reasons for favoring their preferred model over the equivalent versions (Kline, 2015). By considering equivalent models, researchers enhance the reliability and robustness of their study (Kline, 2015).

Physics Identity Framework

Prior studies have shown that physics identity is a crucial motivational factor for explaining students' participation in physics related careers (Godwin et al., 2016; Hazari, Sonnert, Sadler, & Shanahan, 2010). Physics identity pertains to students' perception of themselves as "physics people" and influences their career decisions and academic goals (Hazari et al., 2010). Prior studies have identified three interrelated dimensions of physics identity: perceived recognition by others as a physics person, physics self-efficacy and interest (Hazari et al., 2010). These dimensions have been shown to be important predictors of students' overall physics identity (Godwin et al., 2016; Li & Singh, 2022b).

Perceived recognition in a domain, such as physics, refers to students' perception about whether other people see them as a physics person (Hazari & Cass, 2018). Prior studies have shown that perceived recognition is the strongest predictor of students' physics identity compared to self-efficacy and interest (Cwik & Singh, 2022a; Godwin et al., 2016; Kalender Marshman, Schunn, Nokes-Malach, & Singh, 2019; Li & Singh, 2022a). Moreover, perceived recognition also predicts students' course grades in introductory physics courses (Cwik & Singh, 2022b; Li, Whitcomb, & Singh al., 2020).

Self-efficacy, defined as students' beliefs in their capability to succeed in a certain situation, task, or particular domain (Bandura, 1994; Cwik & Singh, 2023; Kalender et al., 2019; Vincent-Ruz & Schunn, 2017), can influence students' engagement and performance in a given domain (Schunk & Pajares, 2002; Zimmerman, 2000). Students with high self-efficacy often enroll in more challenging courses than those with low self-efficacy because they perceive difficult tasks as challenges rather than threats (Watt, 2006).

Interest is defined by positive emotions accompanied by curiosity and engagement in particular content (Hidi, 2006). Interest has also been shown to influence students' learning outcomes (Hidi, 2006; Smith, Sansone, & White, 2007; Zimmerman, 2000). For example, one study showed that making science courses more relevant to students' lives and transforming curricula to promote interest in learning can improve students' achievement (Häussler & Hoffmann, 2002).

Research suggests that these three motivational constructs correlate to and interact with each other (Bandura, 1994; Hidi & Renninger, 2006), but the predictive relationships among them are unclear. Prior studies have proposed different models to explain these relationships. For example, some prior studies used a model in which self-efficacy is the predictor of both interest and perceived recognition (Godwin, Potvin, Hazari, & Lock, 2016; Lock, Hazari, & Potvin, 2019), while another study used the model in which interest is the predictor of both self-efficacy and perceived recognition (Hazari, Chari, Potvin, & Brewe, 2020). Although most of these studies have presented theoretical frameworks for their models, they have not explicitly discussed the existence of statistically equivalent models and how their proposed models are favorable compared to the equivalent ones based on evidence beyond model fit indices (Crockett, 2012).

In this study, we propose a SEM model in which perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest. This model draws inspiration from our previous qualitative research (Doucette, Clark & Singh, 2020; Doucette & Singh, 2020; Li & Singh, 2023; Santana & Singh, 2021) and other researchers' studies (Bandura, 1994; Hidi & Renninger, 2006). For example, prior research has shown that self-efficacy is influenced by encouragement or discouragement related to one's performance or ability (Bandura, 1994). Similarly, external encouragement is crucial for students to maintain and develop their interest, particularly when faced with difficulties (Bloom, 1985; Hidi & Renninger, 2006). In our prior interviews with students in physics courses, we found that compared to men, women were less likely to feel positively recognized by instructors and teaching assistants (TAs), which adversely affected their self-efficacy and interest (Doucette & Singh, 2020; Li & Singh, 2023). Furthermore, prior studies have shown that people's self-efficacy can influence their interest in a task (Silvia, 2001, 2003). These findings have informed our specification of the SEM model in Figure 1. In this model, perceived recognition predicts self-efficacy and interest, self-efficacy predicts interest, and these three constructs mediate the relation between gender and physics identity (note that although there can be a direct path from gender to identity, we never find that path to be statistically significant in any models examined in this paper, so we have omitted it for clarity). In this study, we test this proposed model and also consider its statistically equivalent models and their potential instructional implications. Additionally, we discuss how intervention studies can further contribute to identifying a more accurate causal model among the statistically equivalent ones.

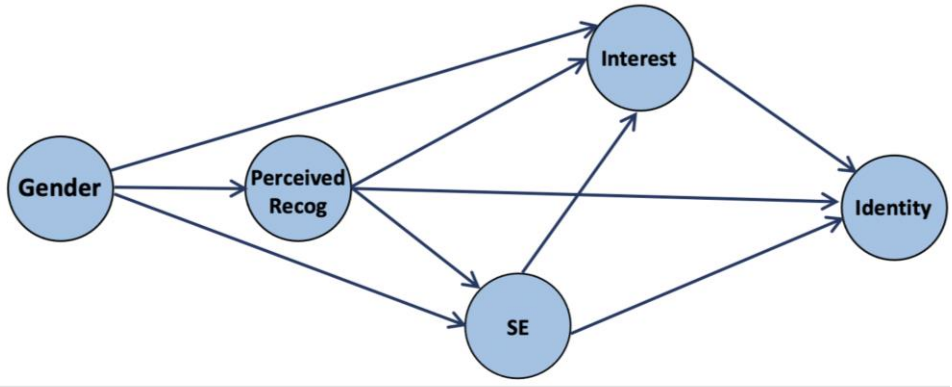

Figure 1. Schematic representation of the proposed SEM model in which perceived recognition (Recog) predicts self-efficacy (SE) and interest, and self-efficacy predicts interest.

Research questions

Our research questions are as follows:

RQ1. What are the gender differences in students' self-efficacy, interest, perceived recognition, and overall physics identity at the end of physics courses?

RQ2. How well does our proposed model fit the data?

RQ3. How many statistically equivalent models does our proposed model have?

RQ4. What are the potential instructional implications of different statistically equivalent models and how can one further determine a more accurate causal model?

Methodology

Participants

The survey data used in this study were collected at the end of the second course of a two-term college calculus-based introductory physics sequence (physics 2) in two consecutive spring semesters. Physics 2 includes topics such as electricity and magnetism, interference, and diffraction. This course is a traditional lecture-based course (4 hours per week) with recitations (1 hour per week) in which students typically work collaboratively on physics problems. This course is generally mandatory and taken by those intending to major in engineering, physical science, and mathematics in the second semester of the first year of their undergraduate studies. In physics 2, after students have been on campus for a semester, the feeling of uncertainty and anxiety during the transition to college may decrease. As students gradually adapt to the new environment, their motivational beliefs may also become more stable. In addition, students in the introductory physics course are admitted to engineering school and school of arts and sciences as undecided majors, and they usually declare their majors in their second year, which is after physics 2. Thus, their physics motivational beliefs at the end of physics 2 are important in influencing their major decisions in engineering and physical science disciplines. Furthermore, the introductory physics sequence is generally mandatory for these students because it is the foundation of many disciplines and contributes directly to engineering, and most scientific fields, so their motivational beliefs at the end of this sequence may also have a long-term influence on their future studies and career. Therefore, in this study, we focus on students' motivational beliefs and the predictive relationships among them at the end of physics 2.

This research protocol was approved and carried out in accordance with the principles outlined in the university institutional review board (IRB) ethical policy. The paper surveys were handed out and collected by TAs in the last recitation class of a semester. We encouraged the instructors to give students small amount of course credit or extra credit for completing the survey. While we did not track the specific implementation of this incentive, most instructors complied with the recommendation. The demographic data of students—such as gender—were provided by the university. Students' names and IDs were de-identified by an honest broker, so researchers could analyze students' data without having access to students' identifying information. There were 915 students participating in this survey, including both semesters. In our final data analysis, we focused on 907 students (including 299 women and 608 men) who provided their gender information. Less than 1% of the participants who did not provide this information were not included in this analysis.

Survey

In this study, our analysis includes four motivational constructs—physics self-efficacy, interest, perceived recognition, and identity. Table 1 shows the survey items for each motivational construct. In our survey, each item was scored on a 4-point Likert scale (1-4) (Likert, 1932) with higher scores indicating greater levels of self-efficacy, interest, perceived recognition, and identity. Most of the self-efficacy and interest items had response options 'NO!, no, yes, YES!', which have been shown to have good psychometric properties and a low cognitive load while reading (Learning Activation Lab, 2017; Vincent-Ruz & Schunn, 2017). The items under physics identity and perceived recognition all had response options 'strongly disagree, disagree, agree, and strongly agree'.

These survey items were adapted from existing motivational research (Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011; Hazari et al., 2010; Learning Activation Lab, 2017; Schell & Lukoff, 2010), and we re-validated them in our prior work (Kalender, Marshman, Nokes-Malach, Schunn, & Singh, 2017; Kalender, Marshman, Schunn, Nokes-Malach, & Singh, 2018; Marshman, Kalender, Schunn, Nokes-Malach, & Singh, 2018; Nokes-Malach, Marshman, Kalender, Schunn, & Singh, 2017; Nokes-Malach, Kalender, Marshman, Schunn, & Singh, 2018). The validation and refinement of the survey involved use of individual interviews with students, exploratory and confirmatory factor analyses (EFA and CFA) (Thompson, 2004), Cronbach's alpha, and Pearson correlation between different constructs (Cronbach, 1951; Hooper et al., 2007; Pearson & Galton, 1895). Given these survey items, the degrees of freedom for our proposed model are 58, which means that our model is over-identified, with sufficient degrees of freedom to estimate the model parameters (Kline, 2015).

Table 1 presents the results of the confirmatory factor analyses, Cronbach's alphas, and fractions of variance explained by each item's corresponding construct. Table 1 shows that all Cronbach's alphas are larger than 0.8 which indicate a high consistency among the items within each construct Moreover, the constructs account for more than 50% (0.5) of the item's variance, indicating that they extract sufficient variance from the items (Kline, 2015). These results provide quantitative support for dividing the constructs as proposed.

Table 2 presents the Pearson correlation coefficients pairwise between the constructs. These constructs have strong correlations with each other, but the correlations are not so high that these constructs cannot be examined separately (Akoglu, 2018). The highest correlation is between physics identity and perceived recognition, consistent with Godwin et al.'s and Kalender et al.'s prior research finding that perceived recognition (external identity) is the strongest predictor of physics identity (internal identity) (Godwin et al., 2016; Kalender et al., 2019).

Table 1. Survey questions for each motivational construct, along with CFA item loadings. Lambda represents the factor loading of each item, which is the correlation between the item and the construct. The square of Lambda for each item gives the fraction of its variance explained by the construct. All Lambdas shown in this table are statistically significant with *p* value < 0.001. †The response options for this question are 'Never, Once a month, Once a week, Every day'. ‡The response options for this question are 'very boring, boring, interesting, very interesting'.

Construct and Item	λ	λ^2
Physics Identity		
I see myself as a physics person.	1.000	1.000
Physics Self-Efficacy (Cronbach's alpha = 0.81)		
I am able to help my classmates with physics in the laboratory or in recitation.	0.698	0.487
I understand concepts I have studied in physics.	0.742	0.551
If I study, I will do well on a physics test.	0.735	0.540
If I encounter a setback in a physics exam, I can overcome it.	0.715	0.511
Physics Interest (Cronbach's alpha = 0.85)		
I wonder about how physics works †	0.701	0.491
In general, I find physics ‡	0.811	0.658
I want to know everything I can about physics.	0.829	0.687
I am curious about recent physics discoveries.	0.720	0.518
Physics Perceived Recognition (Cronbach's alpha = 0.87)		
My family sees me as a physics person.	0.885	0.783
My friends see me as a physics person.	0.908	0.824
My physics TA and/or instructor sees me as a physics person.	0.730	0.533

Table 2. Pearson correlation coefficients of the motivational constructs.

Variable	Physics identity	Self-efficacy	Interest	Perceived recognition
Physics identity	--	--	--	--
Self-efficacy	0.71	--	--	--
Interest	0.69	0.63	--	--
Perceived Recognition	0.83	0.70	0.68	--

Analysis of Survey Data

To analyze the survey data, several statistical techniques were employed. First, we calculated the mean score for each construct for each student. Then, we used a t-test (Cohen, 2013; Gosset, 1908) to compare the mean scores by gender. Before testing the SEM model in figure 1, we first tested for measurement invariance (i.e., whether the factor loadings, intercepts, and residual variances of the observed variables are equal across gender). Then, we performed multi-group SEM analysis to examine whether regression pathways were equal across gender. Results showed that measurement invariance holds and there is no difference in any regression coefficients by gender, which allowed us to use the SEM model in Fig. 1 (Garson, 2014; Kline, 2015).

Results and Discussion

Gender Differences in Motivational Beliefs

Pertaining to RQ1, Table 3 shows that women have significantly lower average scores in all four motivational constructs than men. The effect sizes given by Cohen's d are all in the medium range (Cohen, 2013).

Table 3. Descriptive statistics for women and men's motivational beliefs. M stands for construct mean value, SD is the standard deviation and N is the number of students. All effect sizes (Cohen's *d* values) are statistically significant with $p < 0.001$. A minus sign for Cohen's d indicates that men have a higher construct score than women.

Construct	Women N = 299		Men N = 608		Cohen's *d*
	M	*SD*	*M*	*SD*	
Physics Identity	2.13	0.83	2.67	0.86	-0.63
Perceived Recognition	2.24	0.71	2.71	0.70	-0.66
Self-efficacy	2.64	0.57	2.92	0.56	-0.48
Interest	2.61	0.66	3.02	0.62	-0.65

Estimation of the proposed SEM model

Pertaining to RQ2, we estimate how the proposed SEM model (Model 1) in Figure 1 fit the data. Fig. 2 shows the results of the SEM model. The level of SEM model fit is represented by the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residuals (SRMR), and CFI > 0.9, TLI > 0.9, RMSEA < 0.08, and SRMR < 0.08 are considered as acceptable (Hooper et al., 2007). The model in Fig. 2 fits the data well with CFI = 0.968, TLI = 0.956, RMSEA = 0.063 and SRMR = 0.034 (Hooper et al., 2007). The solid lines represent regression paths, and numbers on the lines are regression coefficients (β values), which represent the strength of regression relations. A regression coefficient reflects the change in the dependent variable (outcome) associated with a one-standard-deviation increase in the independent variable (predictor), while holding other variables in the model constant (Grace & Bollen, 2005).

As shown in Figure 2, gender directly or indirectly predicts perceived recognition, interest, and self-efficacy, which is consistent with the descriptive statistics in Table 3, showing that women had statistically significantly lower score on these three constructs. Figure 2 also shows that perceived recognition, interest, and self-efficacy are all significant predictors of physics identity and perceived recognition is the strongest predictor (with $\beta = 0.56$), which is also consistent with prior studies by Godwin et al. (2016) and Kalender et al. (2019) showing that how students perceive themselves as a physics person is significantly influenced by their perception of how others view them as a physics person. Moreover, Figure 2 shows that perceived recognition also indirectly predicts physics identity through the mediating effects of self-efficacy and interest. Therefore, the total effect of perceived recognition on physics identity is $\beta = 0.56 + 0.44 \times 0.17 + 0.69 \times 0.21 + 0.69 \times 0.29 \times 0.17 = 0.81$. We note that although Table 3 shows that there is a significant gender difference in physics identity, gender does not directly predict physics identity, which indicates that the gender difference is mediated through perceived recognition, interest and self-efficacy.

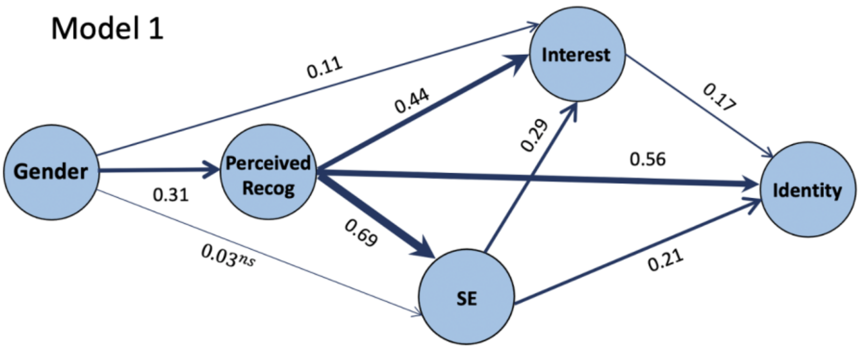

Figure 2. Results of the proposed SEM Model. Each regression line thickness qualitatively corresponds to the magnitude of β values with $p > 0.05$ are indicated by ns (not significant). All β values shown are significant with $p < 0.001$.

Equivalent SEM Models

Our proposed SEM model demonstrates a strong fit with the data. As mentioned earlier, it is important to consider statistically equivalent models that provide alternative representations of the data. Pertaining to RQ3, we found that there are 27 statistically equivalent models in which the relation between gender and physics identity is mediated by self-efficacy, interest, and perceived recognition when we consider the associations among these three mediating constructs. There are three possible associations between each pair. These associations are covariance, direct effect via regression from one to the other, or direct effect via regression in the reverse direction. For example, there can be a direct regression path from self-efficacy to interest or from interest to self-efficacy, or there may only be a covariance between self-efficacy and interest. Similarly, there are three possible types of associations between self-efficacy and perceived recognition, and between interest and perceived recognition. Thus, with the constraints that no regression arrows point to gender and arrows can only point to physics identity since it is the outcome variable, there are $3 \times 3 \times 3 = 27$ statistically equivalent SEM models in total. All 27 models have the same fit indices as our proposed model: CFI = 0.968 (> 0.90), TLI = 0.956 (> 0.90), RMSEA = 0.063 (< 0.08) and SRMR = 0.034 (< 0.08) (Hooper et al., 2007). Thus, these statistically equivalent SEM models are all robust from a statistical point of view.

Instructional implications of the statistically equivalent models

Although all 27 models fit the data equally well, they may have different potential instructional implications. Here, we compare our proposed model with two other statistically equivalent models. In Model 2, self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition. In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition. We focus on discussing these two equivalent models because they are representative of the models in prior research (Godwin et al., 2016; Hazari et al., 2020; Lock et al., 2019) and differ from our proposed model by having self-efficacy and interest predicting perceived recognition.

Models 2 and 3 emphasize that students' perceived recognition is predicted by their interest and self-efficacy. Thus, these models may be interpreted to imply that women are not feeling positively recognized by their instructors and teaching assistants (TAs) as much as men because they have lower interest in physics and lower self-efficacy than men. These implications are different compared with the implication of Model 1, which emphasizes the importance of recognizing students appropriately, which can help students develop their self-efficacy and interest in physics.

We note that Models 2 and 3, in which perceived recognition is predicted by self-efficacy and interest, are not supported by our prior interview findings (Doucette & Singh, 2020; Li & Singh, 2023) showing that perceived recognition may be an important driver of the other two constructs. Our interviews show that women were less likely than men to feel positively recognized by physics instructors/TAs, and this lack of recognition or discouraging feedback from instructors/TAs deteriorated their self-efficacy, and lowered perceived recognition and self-efficacy further lowered their interest (Doucette et al., 2020; Doucette & Singh, 2020; Li et al., 2020; Santana & Singh, 2021). This experience made the female students wonder whether it was because their questions were not good or too easy, and thus they started doubting their ability to do well in their courses. Therefore, Model 1 is more aligned with our interview findings.

Moreover, in Models 2 and 3, gender is the only predictor of interest or self-efficacy which together with the data in Table 3 (which shows gender differences in all motivational constructs studied) can potentially be interpreted from a deficit perspective. For example, Model 3 has the potential to reinforce instructors' fixed mindset about students. While interest can be viewed as modifiable and there is supporting research on this modifiability (Häussler & Hoffmann, 2002), interest-based accounts of gender differences are historically interpreted as fixed and thus may support lack of action by college instructors (Canning, Muenks, Green, & Murphy, 2019; Canning, Ozier, Williams AlRasheed, & Murphy, 2022). That is, if college instructors assume that students' interest predicts perceived recognition and interest is fixed, they may not make effort to recognize students with a focus on closing demographic gaps.

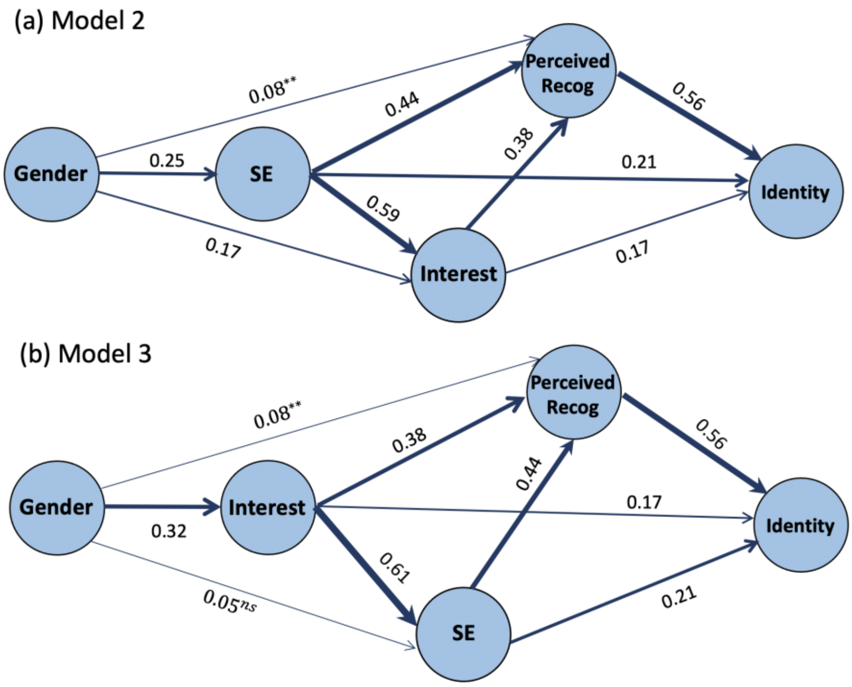

Figure 3. Results of Models 2 and 3. Regression coefficients with $0.001 \leq p < 0.1$ are indicated by ** and with $p > 0.05$ are indicated by ns. All other regression lines show relations with $p < 0.001$.

In this study, we focused on comparing our proposed model (Model 1) with two statistically equivalent models. We note that there are 24 additional equivalent models that have not been discussed. While our interview data provided support for Model 1, we acknowledge that further investigation can gather additional evidence to pinpoint a more accurate causal model. One potential direction for further investigation is to conduct intervention studies. These studies can, e.g., involve instructors implementing interventions that aim to deliberately and effectively recognize students. By doing so, one can assess whether these interventions lead to improvements not only in students' perceived recognition but also in their self-efficacy and interest. If the interventions result in positive changes in all three variables, it would provide further support for our proposed model. Similarly, interventions can also be designed to target students' interest by incorporating interactive activities or introducing engaging topics. By implementing such interventions, one can evaluate whether they have a positive impact not only on students' interest but also on their self-efficacy and perceived recognition.

Summary

In this study, we investigated the predictive relationships among the three dimensions of physics identity: perceived recognition, self-efficacy, and interest. Our results revealed that women scored significantly lower than men in all four motivational constructs, and the gender difference in physics identity is mediated through the gender differences in the other three motivational constructs. Inspired by our prior interviews with students (Doucette & Singh, 2020; Li & Singh, 2023) and previous studies by other researchers (Bandura, 1994; Hidi & Renninger, 2006), we proposed a SEM model to describe the predictive relationships among the constructs studied. Our analysis revealed the existence of 27 statistically equivalent SEM models, all exhibiting identical fit indices but diverging in their predictive relations among the mediating constructs, leading to different potential instructional implications. Our proposed model with perceived recognition predicting self-efficacy and interest emphasizes the important role played by instructor recognition, while other models may inadvertently strengthen instructor fixed mindset about students' interest and self-efficacy (Canning et al., 2019; Dweck, 2008). As researchers, we should consider the statically equivalent models when using SEM and use other evidence from studies such as interviews and intervention studies to support the proposed models or to determine a more accurate causal relationships between the constructs studied.

Limitations and future directions

In this study, we emphasize the importance of considering statistically equivalent models and using extra evidence to determine a more accurate causal model. We discussed two SEM model that are statistically equivalent to our proposed model. There are other statistically equivalent models that have not been discussed in this paper, which would be worth investigating in future studies. In addition, similar to prior studies (Godwin et al., 2013, 2016; Lock et al., 2019), the data used in this study were collected at one time point. Future studies using longitudinal data may provide further understanding of the predictive relationships among students' motivational beliefs.

In this study, since the gender data were collected using only binary categories, we did not have the gender information of students who did not identify as men or women. This issue has been resolved recently in the manner in which the university is now collecting data. However, since the sample size of these students is small (less that 1% of participants), we would not be able to analyze them as separate groups using multigroup analysis in SEM even if we knew their gender identity. Future studies can use other research methods to investigate motivational beliefs of students in other gender categories. In future studies, we also intend to investigate motivational beliefs of students from other underrepresented groups such as ethnic/racial minority students.

This study examined an introductory calculus-based physics course. It would be valuable to investigate the relationship among women and men's motivational beliefs in other physics courses, e.g., for physics majors. Similar studies in different types of institutions and in other countries would also be helpful for developing a deeper understanding of the relationships among students' motivational beliefs in different contexts.

Acknowledgements

This work was supported by Grant No. DUE-1524575 from the National Science Foundation. We would like to thank all students whose data were analyzed and Dr. Robert P. Devaty for his constructive feedback on the manuscript.

References

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91-93.

Bandura, A. (1994). Self-efficacy. In R. J. Corsini (Ed.), *Encyclopedia of Psychology* (2nd ed., Vol. 3, pp. 368-369). Wiley. https://doi.org/10.1002/9780470479216.corpsy0836

Bloom, B. (1985). *Developing talent in young people*. BoD–Books on Demand.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2), eaau4734. https://doi.org/10.1126/sciadv.aau4734

Canning, E. A., Ozier, E., Williams, H. E., AlRasheed, R., & Murphy, M. C. (2022). Professors who signal a fixed mindset about ability undermine women's performance in STEM. *Social Psychological and Personality Science*, *13*(5), 927-937.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Crockett, S. A. (2012). A five-step guide to conducting SEM analysis in counseling research. *Counseling Outcome Research and Evaluation*, *3*(1), 30-47.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cwik, S., & Singh, C. (2022a). Longitudinal analysis of women and men's motivational beliefs in a two-semester introductory physics course sequence for students on the bioscience track. *Physical Review Physics Education Research*, *18*(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.18.020111

Cwik, S., & Singh, C. (2022b). Not feeling recognized as a physics person by instructors and teaching assistants is correlated with female students' lower grades. *Physical Review Physics Education Research*, *18*(1), 010138. https://doi.org/10.1103/PhysRevPhysEducRes.18.010138

Cwik, S., & Singh, C. (2023). Framework for and Review of Research on Assessing and Improving Equity and Inclusion in Undergraduate Physics Learning Environments.

Doucette, D., Clark, R., & Singh, C. (2020). Hermione and the Secretary: How gendered task division in introductory physics labs can disrupt equitable learning. *European Journal of Physics*, *41*(3), 035702. https://doi.org/10.1088/1361-6404/ab7831

Doucette, D., & Singh, C. (2020). Why are there so few women in physics? Reflections on the experiences of two women. *The Physics Teacher*, *58*(5), 297-300. https://doi.org/10.1119/1.5145518

Dweck, C. S. (2008). *Mindset: The New Psychology of Success*. Random House Digital, Inc.

Garson, G. D. (2014). *Structural equation modeling*. G. David Garson and Statistical Publishing Associates.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159-1176. https://doi.org/10.1002/tea.20442

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2013). Understanding engineering identity through structural equation modeling. *2013 IEEE Frontiers in Education Conference (FIE)* (pp. 50-56). IEEE.

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. *Journal of Engineering Education*, *105*(2), 312-340. https://doi.org/10.1002/jee.20118

Gosset, W. S. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1-25. https://doi.org/10.2307/2331554

Grace, J. B., & Bollen, K. A. (2005). Interpreting the results from multiple regression and structural equation models. *Bulletin of the Ecological Society of America*, *84*(4), 283.

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, *39*(9), 870-888. https://doi.org/10.1002/tea.10048

Hazari, Z., & Cass, C. (2018). Towards meaningful physics recognition: What does this recognition actually look like? *The Physics Teacher*, *56*(7), 442-446. https://doi.org/10.1119/1.5055325

Hazari, Z., Chari, D., Potvin, G., & Brewe, E. (2020). The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates. *Journal of Research in Science Teaching*, *57*(10), 1583-1607. https://doi.org/10.1002/tea.21644

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M.-C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching*, *47*(8), 978-1003. https://doi.org/10.1002/tea.20363

Hershberger, S. L., & Marcoulides, G. (2006). The problem of equivalent structural models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 13-41). IAP.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review*, *1*(2), 69-82. https://doi.org/10.1016/j.edurev.2006.09.001

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111-127. https://doi.org/10.1207/s15326985ep4102_4

Hooper, D., Coughlan, J., & Mullen, M. (2007). Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53-60.

Kalender, Z. Y., Marshman, E., Nokes-Malach, T. J., Schunn, C. D., & Singh, C. (2017). Motivational characteristics of underrepresented ethnic and racial minority students in introductory physics courses. *Proceedings of the 2017 Physics Education Research Conference* (pp. 204-207). Cincinnati, OH.

Kalender, Z. Y., Marshman, E., Schunn, C. D., Nokes-Malach, T. J., & Singh, C. (2018). Large gender differences in physics self-efficacy at equal performance levels: A warning sign? *Proceeding of the 2018 Physics Education Research Conference*. Washington, DC. https://www.compadre.org/per/items/4951.pdf

Kalender, Z. Y., Marshman, E., Schunn, C. D., Nokes-Malach, T. J., & Singh, C. (2019). Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way. *Physical Review Physics Education Research*, *15*(2), 020148. https://doi.org/10.1103/PhysRevPhysEducRes.15.020148

Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (Vol. 10). Sage Publications.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. Guilford publications.

Learning Activation Lab. (2017). *Activation lab tools: Measures and data collection instruments*. Retrieved 4 February 2019 from. http://www.activationlab.org/tools/.

Li, Y., & Singh, C. (2021). Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses. *Physical Review Physics Education Research*, *17*(1), 010143. https://doi.org/10.1103/PhysRevPhysEducRes.17.010143

Li, Y., & Singh, C. (2022a). Do female and male students' physics motivational beliefs change in a two-semester introductory physics course sequence? *Physical Review Physics Education Research*, *18*(1), 010142. https://doi.org/10.1103/PhysRevPhysEducRes.18.010142

Li, Y., & Singh, C. (2022b). Inclusive learning environments can improve student learning and motivational beliefs. *Physical Review Physics Education Research*, *18*(2), 020147.

Li, Y., & Singh, C. (2023). The impact of perceived recognition by physics instructors on women's self-efficacy and interest. *arXiv preprint arXiv:2303.07239*.

Li, Y., Whitcomb, K., & Singh, C. (2020). How perception of being recognized or not recognized by instructors as a “physics person” impacts male and female students’ self-efficacy and performance. *The Physics Teacher*, *58*(7), 484-487. https://doi.org/10.1119/10.0002067

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 55. https://legacy.voteview.com/pdf/Likert_1932.pdf

Lock, R. M., Hazari, Z., & Potvin, G. (2019). Impact of out-of-class science and engineering activities on physics identity and career intentions. *Physical Review Physics Education Research*, *15*(2), 020137.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226.

Marshman, E., Kalender, Z. Y., Schunn, C., Nokes-Malach, T., & Singh, C. (2018). A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences. *Canadian Journal of Physics*, *96*(4), 391-405. https://doi.org/10.1139/cjp-2017-018

Nokes-Malach, T., Marshman, E., Kalender, Z. Y., Schunn, C., & Singh, C. (2017). Investigation of male and female students' motivational characteristics throughout an introductory physics course sequence. *Proceedings of the 2017 Physics Education Research Conference* (pp. 276-279). Cincinnati, OH.

Nokes-Malach, T. J., Kalender, Z. Y., Marshman, E., Schunn, C. D., & Singh, C. (2018). Prior preparation and motivational characteristics mediate relations between gender and learning outcomes in introductory physics. *Proceedings of the 2018 Physics Education Research Conference* (pp. https://doi.org/10.1119/perc.2018.pr.Nokes-Malach). Washington, DC. https://www.compadre.org/per/items/4978.pdf

Pearson, K., & Galton, F. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(347-352), 240-242. https://doi.org/doi:10.1098/rspl.1895.0041

Santana, L. M., & Singh, C. (2021). Negative impacts of an unwelcoming physics environment on undergraduate women. *Physics Education Research Conference 2021* (pp. 377-383). Virtual Conference. https://www.compadre.org/per/items/detail.cfm?ID=15784

Schell, J., & Lukoff, B. (2010). Peer instruction self-efficacy instrument [Developed at Harvard University] (unpublished).

Schunk, D. H., & Pajares, F. (2002). The Development of Academic Self-Efficacy. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation: A Volume in the Educational Psychology Series* (pp. 15-31). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012750053-9/50003-6

Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, *5*(3), 270-290.

Silvia, P. J. (2003). Self-efficacy and interest: Experimental studies of optimal incompetence. *Journal of vocational behavior*, *62*(2), 237-249.

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, *99*(1), 99-114. https://doi.org/10.1037/0022-0663.99.1.99

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis*. American Psychological Association.

Vincent-Ruz, P., & Schunn, C. D. (2017). The increasingly important role of science competency beliefs for science learning in girls. *Journal of Research in Science Teaching*, *54*(6), 790-822. https://doi.org/10.1002/tea.21387

Watt, H. M. (2006). The role of motivation in gendered educational and occupational trajectories related to maths. *Educational Research and Evaluation*, *12*(4), 305-322.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82-91.