Supplemental online materials for

Effects of Discovery, Iteration, and Collaboration in Laboratory Courses on Undergraduates' Research Career Intentions Fully Mediated by Student Ownership

Lisa A. Corwin, Christopher R. Runyon, Eman Ghanem, Moriah Sandy, Greg Clark, Gregory C. Palmer, Stuart Reichler, Stacia E. Rodenbusch, Erin L. Dolan[*]

[*]Corresponding author. Email: eldolan@uga.edu

This supplement contains the following:

MEASURES

Note: Items are numbered to allow for easy comparison with factor loading results presented in Table S1.

Discovery scale

In this course I was expected to…	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree
1. generate novel results that are unknown to the instructor and that could be of interest to the broader scientific community or others.						
2. conduct an investigation to find something previously unknown to myself, other students, and the instructor.						
3. formulate my own research question or hypothesis to guide an investigation.						
4. develop new arguments based on data.						
5. explain how my work has resulted in new scientific knowledge.						

Iteration scale

In this course …	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree
1. I was expected to revise or repeat work to account for errors or fix problems.						
2. I had time to change the methods of the investigation if it was not unfolding as predicted.						
3. I had time to share and compare data with other students.						
4. I had time to collect and analyze additional data to address new questions or further test hypotheses that arose during the investigation.						
5. I had time to revise or repeat analyses based on feedback.						
6. I had time to revise drafts of papers or presentations about my investigation based on feedback.						

Collaboration scale

In this course I was encouraged to…	Weekly	Monthly	One or two times	Never
1. discuss elements of my investigation with classmates or instructors.				
2. reflect on what I was learning				
3. contribute my ideas and suggestions during class discussions.				
4. help other students collect or analyze data.				
5. provide constructive criticism to classmates and challenge each other's interpretations.				
6. share the problems I encountered during my investigation and seek input on how to address them.				

Cognitive ownership scale

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. My research will help solve a problem in the world.					
2. My findings were important to the scientific community.					
3. I faced challenges that I managed to overcome in completing my research project.					
4. I was responsible for the outcomes of my research.					
5. The findings of my research project gave me a sense of personal achievement.					
6. I had a personal reason for choosing the research project I worked on.					
7. The research question I worked on was important to me.					
8. In conducting my research project, I actively sought advice and assistance.					
9. My research project was interesting.					
10. My research project was exciting.					

Emotional ownership scale

	Very slightly	Slightly	Moderately	Considerably	Very strongly
1. To what extent does the word delighted describe your experience of the laboratory course?					
2. To what extent does the word happy describe your experience of the laboratory course?					
3. To what extent does the word joyful describe your experience of the laboratory course?					
4. To what extent does the word astonished describe your experience of the laboratory course?					
5. To what extent does the word surprised describe your experience of the laboratory course?					
6. To what extent does the word amazed describe your experience of the laboratory course?					

Intentions to persist in a science research career

To what extent do you intend to pursue a science related research career?

1 2 3 4 5 6 7 8 9 10

Prior to this course, to what extent did you intend to pursue a science related research career?

1 2 3 4 5 6 7 8 9 10

Measurement Model Testing and Model Modifications

Following best practices for structural equation modeling (SEM), we examined the measurement models in isolation prior to using them to specify latent variables in the structural model because misfit in a measurement model can propagate into the overall fit of a full structural model. This can result in misinterpretations about the hypothesized relationships between latent variables based on structural model fit (or lack thereof), when in fact the misfit is within the measurement model. It is important to take the nested nature of the data into account when using CFA or SEM with clustered data. Failing to do so has the potential for the analyses to result in improper fit statistics, biased parameter estimates, and attenuated standard error estimates, much like traditional regression analyses when the clustering of data is ignored (Pornprasertmanit, Lee, & Preacher, 2014; Wu & Kwok, 2012). Because our research questions concerned within-student relations of the variables, we only needed to make statistical adjustments to the model (as described in the section below on Structural Model Testing**)** rather than conduct a full two-level CFA or SEM (Stapleton, Yang, & Hancock, 2016). Thus, we performed all CFA and SEM analyses using the R software for statistical computing (R Core Team, 2016), using the 'lavaan' (Rosseel, 2012) and 'lavaan.survey'(Oberski, 2014) packages. We used robust maximum-likelihood (MLR) estimation to account for any departures from normality in the data. We used multiple imputation to account for missing data in the analyses, multiply-imputing 100 datasets using the 'mice' package (Buuren & Groothuis-Oudshoorn, 2011). The lavaan.survey package allows for use of a multiply-imputed dataset to be used in the desired analyses (including the statistical correction to account for the nested data), combining the results of the analyses on each imputed data set via Rubin's rules (Rubin, 2004). Full details on using multiply-imputed data sets for complex survey analyses of structural equation models can be found in Oberski (2014). We also generated factor scores for the cross-course comparisons (detailed in the primary article) using the 'lavaan' package.

We tested the measurement models for discovery, iteration, collaboration, and ownership (i.e., latent variables) using confirmatory factor analysis (CFA; see Table S1 for factor loadings for each scale). The CFA on the iteration subscale indicated good model fit (Table S2). However, the models for both the discovery and collaboration subscales had fit statistics that indicated slight model misfit. In both cases the TLI was lower than the traditional cutoff of 0.90 and the RMSEA was greater than 0.10. These are indicative of unmodeled relationships in the data (i.e., misspecified factor loadings). In the case of the one-factor models, it indicates that some indicators may also be "loading" on an unmodeled factor, which causes the relationships between some indicators to be greater than the general relationship between the group of indicators. In our specific case, the modification indices suggested that the error of items 4 and 5 in the Collaboration scale should be correlated and the errors between items 3 and 4 in the Discovery scale should be correlated. This indicates that something is making these items correlate above and beyond the shared variance by the other items on the scale. We opted to correlate the errors because they were on the same subscale (i.e., Collaboration), so it would not change what the latent variable was contributing to the model (subscale was still unidimensional). Also, we wanted to assess the relationships between the latent variables in the structural model rather than pursue modifications to existing measures.

The CFA on both the cognitive ownership and emotional ownership subscales indicated that substantial revisions were necessary as the fit indices were much lower than the traditional cutoff values. The modification indices for the cognitive ownership subscale suggested correlating the errors for several items: 1 with 2, 6 with 7, and 9 with 10. When examining these pairs of variables, it seems reasonable that the content, wording, or juxtaposition of the items caused these pairs to be correlated beyond the shared correlation between the items as a whole. The modification indices for the emotional ownership subscale suggested that the last three items (14, 15, 16) correlated above and beyond the shared factor correlation. This suggests that the scale may better be explained by dividing these items into two factors, one representing enjoyment (11, 12, 13) and the other representing surprise (14, 15, 16). Because our aim was to understand the relationships between the latent variables in the model, we opted not to examine the measurement model further. Instead, we correlated the errors for the three pairs of items on the cognitive ownership scale and the last 3 items on the emotional ownership subscale to capture additional relationship between these items. Thus, for the present models, emotional ownership represents a blend of enjoyment and surprise; for our study, it was not necessary to understand any nuanced differences between these emotional states.

Table S1. Standardized Item Loadings and Standard Errors for Measurement Model Testing					
Scale	Laboratory Course Assessment Scale			Project Ownership Survey	
Subscale	Iteration	Discovery / Relevance	Collaboration	Cognitive	Emotional
Item 1	.68 (0.03)	.85 (0.02)	.63(0.04)	.64 (0.02)	.93 (0.01)
Item 2	.81 (0.02)	.83 (0.02)	.62 (0.04)	.68 (0.03)	.92 (0.01)
Item 3	.72 (0.03)	.67 (0.03)	.67 (0.09)	.68 (0.02)	.95 (0.01)
Item 4	.84 (0.01)	.76 (0.02)	.56 (0.04)	.63 (0.02)	.72 (0.03)
Item 5	.88 (0.01)	.85 (0.02)	.67 (0.02)	.79 (0.01)	.64 (0.04)
Item 6	.73 (0.03)	--	.66 (0.04)	.63 (0.03)	.80 (0.01)
Item 7	--	--	--	.77 (0.01)	--
Item 8	--	--	--	.67 (0.03)	--
Item 9	--	--	--	.81 (0.01)	--
Item 10	--	--	--	.81 (0.02)	--

Note: Item numbers correspond with the order they are presented in the supplemental materials. The final loadings and standard errors are provided for those model that included modifications.

Table S2. Measurement Model Fit and Subscale Reliability

Subscale	df	Robust Chi-Square	CFI	TLI	RMSEA	SRMR	Alpha
Iteration	9	43.34	0.98	0.96	0.10	0.03	0.90
Discovery	5	145.51	0.92	0.84	0.22	0.04	0.90
Discovery*	4	81.93	0.97	0.92	0.16	0.02	--
Collaboration	9	41.45	0.93	0.89	0.12	0.04	0.79
Collaboration*	8	25.71	0.97	0.94	0.08	0.03	--
Ownership (2-Factor)	103	1527.51	0.84	0.82	0.141	0.06	--
Cognitive Ownership	35	647.50	0.82	0.77	0.18	0.07	0.91
Cognitive Ownership*	32	235.84	0.95	0.93	0.10	0.04	--
Emotional Ownership	9	504.189	0.87	0.78	0.30	0.07	0.94
Emotional Ownership*	2	1.87	1.00	1.00	0.00	0.00	--

*Model fit with modification indices. Alpha is not reported a second time for the models, as the reliability of the items is independent of model structural fit and thus would be equivalent.

Structural Model Testing

SEM uses factor scores when estimating structural paths (i.e., relationships between variables) because factor scores account for the unique unreliability for each item on a scale. Specifically, factor scores include adjustments of observed scores for each item (e.g., mean of responses to that item) by the strength of the individual factor loading for the item. This is distinct from traditional summed or average scores, which give each item equal weight and thus implicitly assume that all items are perfectly reliable (i.e., they have the same factor loadings or all perfectly reflect the factor or latent variable). By weighting the observed scores by the reliability of the items (i.e., factor loadings), we were able to focus on how students differed on their levels of the latent variables (e.g., sense of ownership), even if they had the same observed responses (e.g., same mean or sum of responses on the ownership scale). In other words, by using factor scores, we gave preference to items that better measured the latent variable.

To assess model fit for all CFA and SEM models we examined the chi-square test of model fit, the root-mean-square-error-of-approximation (RMSEA; (Steiger, 1990), the standardized root mean-square residual (SRMR; (Joreskog & Sorbom, 1981), the Tucker-Lewis Index (TLI; (Tucker & Lewis, 1973), and the comparative fit index (CFI; (Bentler, 1990). The first three indices are absolute fit indices, indicating how well the data reproduce the implied model. The last two indices are relative fit indices, comparing the fit of the model to a null model where no relationships are posited between any of the variables. Low values of the SRMR and RMSEA and high values of the TLI and CFI imply that hypothesized models are plausible explanations of the data (Hu & Bentler, 1998; 1999).

We tested the initial full structural model with the model specifications as suggested by the CFAs, including the five error correlations, and this model adequately fit the data. The chi-square test was significant (as is expected when using CFA/SEM with large sample sizes; Kline, 2015), but the CFI and TLI were close to the traditional cutoff values (0.941 and 0.935, respectively). The RMSEA indicated good model fit (0.052), as did the SRMR (0.044). We encourage the interested reader to consult Bandalos & Finney (2010)or Schreiber, Nora, Stage, Barlow, & King (2006)for more information regarding model fit statistics, their acceptable ranges, and potential other measures of model fit. Importantly, because the model had adequate fit, we could meaningfully interpret the structural paths in the model and test our overarching hypothesis.

Table S3. SEM Model Fit. Models 1A and B and 2A and B are depicted in Figure 1 of the main manuscript. Models 1C and D and 2C and D are included on the following page.

	R^2	df	Robust Chi-Square	CFI	TLI	RMSEA	SRMR
Model 1A: Discovery, Iteration, Collaboration → Ownership → Intentions (*indirect effects*)	0.11	508	1566.171	0.94	0.93	0.05	0.05
Model 1B: Discovery, Iteration, Collaboration → Ownership → Intentions (*direct and indirect effects*)	0.11	505	1560.51	0.94	0.93	0.05	0.05
Model 1C: Discovery, Iteration, Collaboration → Intentions	0.04	128	434.679	0.94	0.93	0.05	0.04
Model 1D: Ownership → Intentions	0.11	111	495.254	0.96	0.96	0.07	0.04
Model 2A: Model 1A plus Previous Intentions	0.45	541	1644.26	0.94	0.93	0.05	0.05
Model 2B: Model 1B plus Previous Intentions	0.45	538	1634.04	0.94	0.93	0.05	0.05
Model 2C: Model 1C plus Previous Intentions	0.43	145	476.83	0.97	0.96	0.06	0.04
Model 2D: Model 1D plus Previous Intentions	0.45	127	513.862	0.96	0.96	0.06	0.05

Figure S1. Additional models tested. We tested four additional structural models (described in Table S3) to gain more insight into the relationships among course design features, student ownership, and students' career intentions, which we present here for comparison. All depicted relationships are significant. Relationships between predictors and outcomes are in black and correlations among predictors and between mediators are in blue.

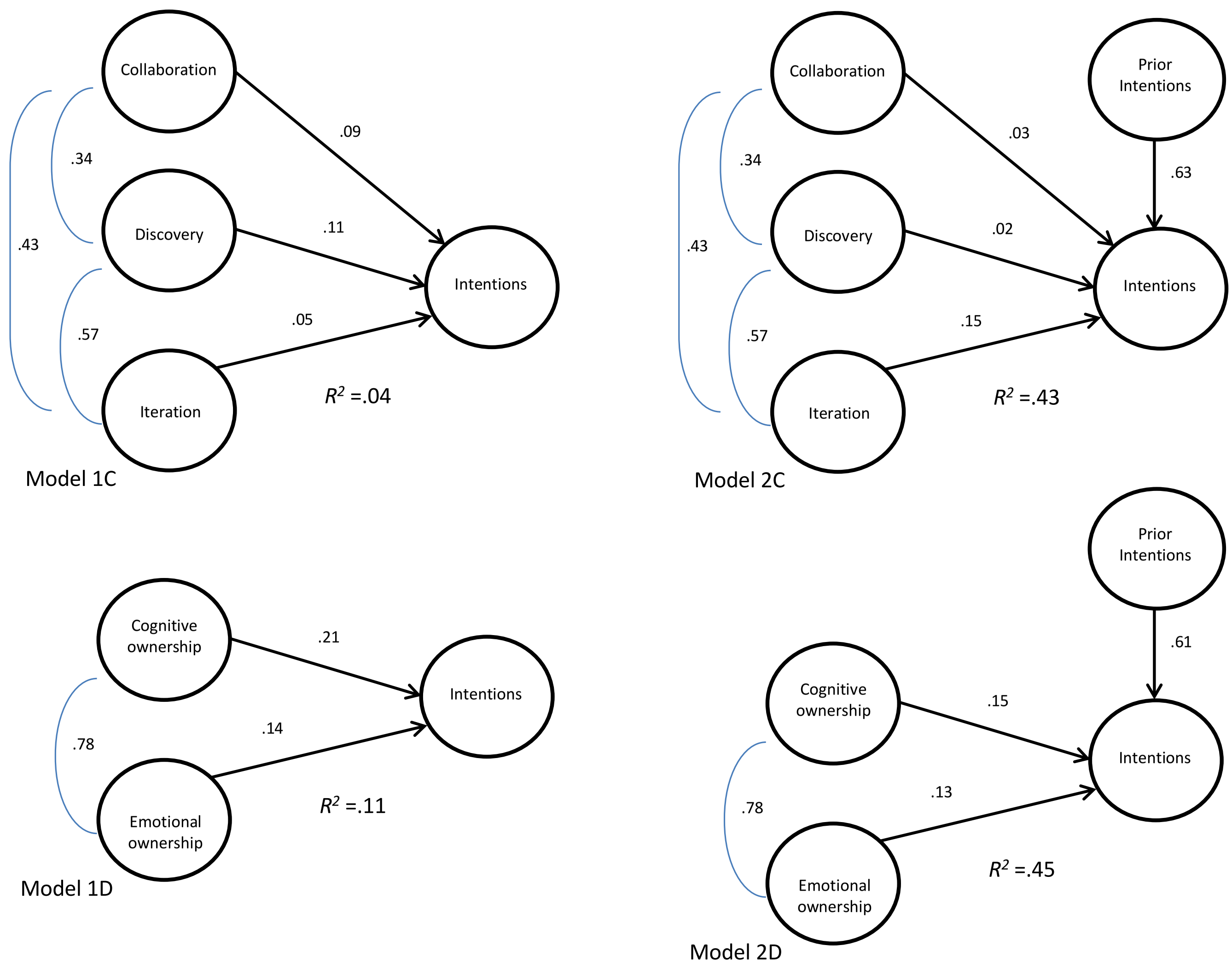

Table S4. Pre- versus post-course intentions to pursue a science research related career

	Pre-course intent									
Post-course intent	1	2	3	4	5	6	7	8	9	10
1	20	5	4	2	2	1	0	2	0	0
2	8	9	6	6	2	2	2	0	0	1
3	12	8	15	3	9	1	5	2	2	2
4	6	3	6	6	10	6	3	1	0	0
5	1	5	9	9	24	10	4	2	0	3
6	3	4	10	8	15	15	10	5	1	6
7	3	5	2	12	17	17	34	5	2	6
8	0	3	0	1	4	14	7	11	3	0
9	0	0	0	1	3	6	6	8	15	2
10	2	1	6	13	8	11	11	9	9	107

We used post-course research career intentions as the outcome in fitting our structural models. Here we provide students' raw pre- and post-course career intentions (N=680) to illustrate that ~38% of students (n=256) do not shift in their intentions (i.e., values along the shaded diagonal). This suggests that our models that include prior intention (Models 2A and 2B in the main manuscript and Models 2C and 2D in the supplement) are likely to be describing change in intention because students who did not change in their intention have their post intentions fully explained by their prior intentions. Of the students whose intentions change, most are increasing (n=286, or 42%). About 16% of our sample is reporting the maximum level of intent prior to their course (i.e., 107 students rated their pre-course intentions as 10), which limits the variance we can observe in their responses.

REFERENCES

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, 93–114.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45.3*.

Joreskog, K. G., & Sorbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods*. Uppsala, Sweden: University of Uppsala, Department of Statistics.

Oberski, D. L. (2014). lavaan. survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, *57.1*, 1–27.

Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring Clustering in Confirmatory Factor Analysis: Some Consequences for Model Fit and Standardized Parameter Estimates. *Multivariate Behavioral Research*, *49*(6), 518–543. https://doi.org/10.1080/00273171.2014.933762

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct Meaning in Multilevel Settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. https://doi.org/10.3102/1076998616646200

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Wu, J.-Y., & Kwok, O. (2012). Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(1), 16–35. https://doi.org/10.1080/10705511.2012.634703