Appendix A: Description of Education Intervention

The approach to creating the course's curriculum was based on five key components: 1. Positioning the students' actions as a point of empowerment rather than a point of blame; 2. Developing a continuous knowledge-action-reflection cycle; 3. Fostering collaboration amongst students from different disciplines; 4. Making global issues, such as climate change, relevant to the students careers and lives; 5. Focusing on higher-order knowledge and subjective ways of knowing (value-laden knowledge). The central tenant our of course was to let the students choose what sustainable actions they wanted to focus on based on what was most relevant to their personal lives and was an appropriate scale for a one semester course. We then connected the research and broader problems and solutions to their selected focus/action, bridging the 'knowledge-action gap' (Brundiers & Wiek, 2013; Frisk & Larson, 2011). Due to the personal nature of individual behavior change, we also built in moments for reflection and positive reinforcement, therefore expanding the traditional knowledge to action cluster to a more robust knowledge-action-reflection cycle. The ultimate goal of this course was to leave the students feeling empowered about their ability to foster change, hopeful about the diverse and exciting ways in which sustainable change can be achieved, and engaged in sustainability as an area of lifelong interest.

A1. EfS Informed Pedagogy: The pedagogy and course design were based on the Education for Sustainability framework initially laid out in Frisk & Larson (2011). Here we briefly explain three key examples of how the implementation of this framework operated:

- **Student Choice** As can be seen in Table A1, internal motivation, early success, and consistency are critical components of fostering pro-environmental behaviors. With relation to self-determination theory, the students had autonomy to choose their focus from the very onset of class in order to position the behavior change as an internal rather than external choice. According to researchers it is important to enhance an internal locus of control because with external controls, the person will revert to pre-intervention behaviors if the external force is removed resulting in negligible long-term change. Focusing on self-selected behaviors also speaks to intrinsic satisfaction because students have the opportunity and support to accomplish their personal behavioral goal, hence achieving

Table A1. Examples of Relevant Theories

Theory	Brief Summary of Theory
Self-determination theory (SDT)	People are more likely to engage in a behavior if they perceive that the motivation to do it comes from within them rather than from an external, controlling agent.
Self-efficacy theory	People that experience success at attaining their goals develop stronger intentions to continue to perform them.
Cognitive Dissonance Theory	People strive for internal consistency, meaning that they want to obtain information that is consistent with their actions and they want to behave consistently across related actions.
Locus of control	People often refer to locus of control as internal—meaning the person feels they have control of events/actions—or external—meaning the person feels external powers have control over events/actions.
Intrinsic satisfaction	People often act for self-centered reasons & when fostering pro-environmental behavior it is critical to speak to people's intrinsic desire—rather than solely altruistic—to complete tasks (DeYoung, 2000).

behavioral competence (De Young, 2000). However, it is critical to ensure that the task selected is feasible so that students have a positive experience with Sustainable behaviors (i.e., feel empowered) and feel the satisfaction of task completion. As noted by De Young (p. 521), "What seems to others a simple action may become for them a major challenge" hence, reinforcing the concept that we, as teachers, cannot be the determiners of feasible behaviors for the students because we may not understand each individual's barriers to change.

Example: Students self-selected the behavior they wanted to focus on for their final project.

- **Social Support and Networks**: Additionally, community-based social marketing (McKenzie-Mohr, 2011) and diffusion theory (Rogers, 2003) provide strong real-world justifications for focusing on locally relevant barriers and strategies based on the participants' attitudes and perspectives rather than externally-selected strategies (i.e., chosen by teachers). Unlike the behavioral theories described above, community-based social marketing (CBSM) and diffusion theory outline methods and procedures for effectively targeting sustainable change based on a number of different theoretical constructs as well as significant number of real-world case studies. CBSM is commonly used to foster sustainable behaviors in developed countries (e.g., Canada, US), whereas diffusion theory has frequently focused on innovations, such as improved cook-stoves, pure water spigots, and photovoltaics, in developing countries. While the contexts and actions targeted differ between these two approaches, both Rogers and McKenzie-Mohr highlight the importance of understanding local barriers and developing strategies for overcoming those barriers with local stakeholders. Rogers and McKenzie-Mohr also advocate for the use of local networks in spreading the uptake of a targeted behavior. Rogers writes that 'individuals depend mainly on the communicated experience of others much like themselves who have already adopted a new idea' (p. 331) and advocates for developing 'diffusion networks.' Similarly, McKenzie-Mohr suggests asking those already engaged in the targeted sustainable behavior to talk to their neighbors and spread their positive experience with their community.

  Examples: During the course, we began by asking the students about their perceived barriers to change and assigned them to develop strategies for overcoming those barriers. We re-visited this topic throughout the semester, particularly as part of the final project and reflection. In order to enable the students to share their actions and commitments with their peers, we hosted a Sustainability fair in which the students created booths with visuals and activities. For example, the students that had committed to composting brought in their composting bins and showed the fair's visitors how to build and maintain their own composting bins. Some groups even utilized our commitment approach by asking visitors to commit to various Sustainability activities. Yet another group, hung photos of professors using reusable water bottles in order to establish their use as the social norm. Other groups created Facebook pages in order to increase the size of their diffusion network while also sharing advice, strategies, and photos of their projects. We posted photos of each student with their project and commitment on the university webpage, making the commitment public, increasing the number of people that are aware of our students' actions, and encouraging students to behave consistently. Through utilizing peer engagement and building social networks focused, the students created a support system for their sustainability-related actions that will be in place even after our interaction with them as teachers comes to an end.

- **Systems Thinking** With a strong focus on 'systems' throughout the course, we aimed to connect students' targeted behaviors with broader issues and an array of actions. We hoped that this would further engage students in the material because they were learning about system interactions that justified their behaviors and they could connect their chosen activity to a range of other behaviors. These connections create a sense of internal consistency as well as ideally foster spillover, hence taking into account cognitive dissonance theory.
  Example: In small groups students drew out a systems diagram of one aspect of the food system (e.g. farming) and then as a class discussed how these parts were connected and interrelated.

A2. Course Outline: We think this course outline is broadly adaptable to different instructor preferences and institutional circumstances for an Introduction to Sustainability.

1. **The Hook:** It is really important when introducing Sustainability to grab the students' attention right away; before dumping information on them or even providing any backstory. For our hook the students completed an ecological footprint calculator and we spent a whole class doing the *Fishbanks: a Renewable Resource Management Simulation* (https://mitsloan.mit.edu/LearningEdge/simulations/fishbanks/Pages/fish-banks.aspx). Both of these activities were very engaging for the students, demonstrated a serious problem which created a need for Sustainability and motivated them to learn more about how we might solve these issues.
2. **Quick Overview:** We gave a very brief overview of the types of problems which drove the creation of Sustainability science, a look at their underlying causes (i.e. population and consumption growth) and how Sustainability science thinks we might go about tackling them. This in total took less than two classes worth of time.
3. **Climate:** The only area we delved into detail was climate change. We think this is such an important and significant problem that every citizen needs to have at least a basic understanding of it.
4. **Solution Spaces:** This was the meat of the course, organized around solutions/actions which the average person can take in their life. We delineated six solution spaces: Food, Waste, Electricity, Water, Transportation and Consumption. For each space we studied the whole system, their role in that system and specific actions which they could take that would improve the Sustainability of the system. Due to time constraints we could only spend significant time on Food and Waste while the others were much more briefly covered.
5. **Personal/Group Action:** For their final project the students worked in teams that were created based upon what solutions space most interested them. The teams were charged with promoting sustainable actions for that space at a fair they put on for the entire university. In addition the students had an individual assignment where they chose a specific action for their life. The action chosen was not important but they had to justify how that action would actually improve the sustainability of their system and clearly describe the steps (procedural) that one would have to take to achieve this more sustainable behavior.

A3. What we didn't do: It is very important to emphasize some of the things we purposely did not do in this course.

1. We did NOT try to cover every sustainability topic. The typical introductory course attempts to cover every main area within that discipline (Intro bio, chemistry, etc.). Sustainability as a field is so enormous that this task is basically absurd. It just doesn't work. It leaves the student overwhelmed and no time in the course to get into depth about anything. Far more important is to get the student interested in Sustainability, arm them with the right analytical skills and motivate them to find out more. We believe motivation for continued learning is strongest when someone is striving to become more Sustainable in their personal life.
2. The course was NOT organized by academic topic: One approach to Sustainability and other interdisciplinary style courses is to organize them by the academic areas they synergize. Have a unit on economics, then anthropology then ecology, etc. The other common alternative is to do it based on problems areas. A unit on pollution, then biodiversity etc. These categorizations are arbitrary (as are all categorizations in our interconnected world) but the real problem is that they aren't meaningful to the average person. Outside of academia what does it matter if an idea came from anthropology or from economics. Problem organization is useful but it is very disempowering. Better to organize based on solutions and work back to the problems.
3. We did NOT focus on facts. Data and facts were mentioned in presentations and videos but we made a very clear point that it was not necessary to memorize them. The goal was to understand systems. So we never tested based on facts and all work was assessed based on an understanding of connections not on any particular fact.

The course was called *Sistemas Socio-ecologicos para la Sostenibilidad* and was run from January 2014 to May 2014 at the *Escuela Nacional de Estudios Superiores* in Leon, Mexico. We had 87 students enroll in the course from four different majors. Unfortunately taking an interdisciplinary class such as this is very challenging for students given the confining schedules of their degree programs. We were only able to find a 2 hour slot once a week and even that didn't work for many interested students. Between vacations and university-wide cancelations we were only able to hold 15 classes, which gave us a mere 30 hours of facetime with the students. On top of that the students estimated that they missed an average of nearly 3 classes due to obligations in their degree program and were late to more than 4 additional ones.

Yet overall the course was seen very positively by the students, 95% of them saying that they "liked" or "liked a lot" the course. All of the survey respondents said they would recommend it to fellow students and all but one would probably or definitely enroll in a follow up course. Most importantly for us 100% agreed that "Due to their participation in the course changed their opinion on the urgency of resolving sustainability challenges."

If you are interested in more details about the course or specific materials please don't hesitate to contact the authors.

Appendix B: Data Collection Methodology

B1. Study Context

This study took place at a new public university in Leon, Guanajuato, Mexico. Leon is the seventh largest city in Mexico with over a million and a half residents. The city has experienced a lot of growth over recent decades, propelled by its nation-leading shoe industry. From a Sustainability perspective the city is known internationally for a system of protected bike lanes and a bus rapid transit system as well as one of the best municipal water systems in Mexico. From a waste perspective though the city lags behind many other Mexican cities. There is no separate pick-up for recycling or organics and according to the director, who we interviewed, a transfer stations at which to sort waste was at least a decade away.

In 2011 Mexico's National University (UNAM) inaugurated its first independent campus expansion outside of the Mexico City area. UNAM is Mexico's premier university, highly selective and sought after for its free tuition and high quality education. The branch in Leon was named *Escuela Nacional de Estudios Superiores* (ENES) and has gotten started with only a handful of degree programs and at the time of the study ~500 students and several dozen faculty. The university draws mostly from Leon and the state of Guanajuato, though students come from all over the country as well. Unlike at UNAM (and most Mexican Universities), neither the departments nor the students are segregated from each other spatially. Yet functionally the degree programs are 100% independent of each other (in terms of class requirements etc).

In 2012 we were invited to establish a Sustainability department at the University. We explored developing an independent degree program but found the best educational opportunity to be in introducing Sustainability to all the students. We took over the teaching of a course in three of the degree programs and offered an elective which was taken by most of the students in the other two degree programs (a sixth and seventh degree have since been started). Having complete control of the elective we were offering, we saw in opportunity to investigate whether an educational impact could have a real and significant impact on behavior. As detailed in the Appendix A we designed the course with this research question in mind.

B2. Study Population

Ideally we would have randomly sampled the student population for our baseline information and randomly sampled for participants in the intervention. This was not feasible. Instead we make the case that our study population is sufficiently representative of ENES's student population. For three of the degree programs we surveyed every student from one cohort. For the other two degree programs the students enrolled voluntarily. There is potentially a bias because the students chose to take a Sustainability elective but we argue that in actuality that bias was very limited. This is because there are almost no electives offered outside of degree programs so ours was effectively the only option whether you were into Sustainability or not. Our understanding was that everyone from these two degree programs which did not enroll in our elective (from the cohorts which were allowed to take electives) did not do so because they had a scheduling conflict with a course in their degree program. Therefore we captured almost all the students from a single cohort in each degree program ENES was offering at the time.

We statistically analyze whether the sub-sample of the intervention and the follow-up can be considered drawn from the same population as our broad sample (described in the previous paragraph) in Appendix D.

B3. Subjective Knowledge

Before beginning this project we had concluded that we would no longer divide subjective knowledge into two separate domains, effectiveness and social knowledge as we had done previously (Redman & Redman, 2014). There were three basic reasons for this:

- There is a clear consensus in the psychological and behavior change literature that subjective knowledge is a key motivator of human behaviors. But there is virtually an endless array of exact theories, models and proposed mechanisms. Evidence has been found in support of many of them but none are convincing as being complete explanations for human behavior. The use of effectiveness knowledge and social knowledge in (Frisk & Larson, 2011) was an attempt to reconcile and synergize these theories. But we now propose that the more coherent synergy is to group all of them under Subjective knowledge.
- From a research perspective designing survey questions which only got at social knowledge or only effectiveness knowledge proved impractical. We ultimately felt that we could not rigorously defend why one question measured effectiveness and another social and not vice versa.
- Most importantly the division of subjective knowledge was actually a hindrance for designing curriculum based on EfS and made the approach more difficult to explain to practitioners. Teachers for example rapidly grasped the concept of procedural knowledge and the overall idea of subjective knowledge, but not the finer distinctions. Rather they were interested in the findings of psychology for subjective knowledge; building social norms in the classroom, consistency between words and actions, focus on values, social networks, locus of control etc. The key for EfS is to acknowledge the central role of subjective knowledge for behavior change and then draw on these specific findings to select the curriculum that you will use.

B4. Instrument Design

The survey instrument was drawn from our previous work in (Redman & Redman, 2014) on food and waste behaviors. We chose to focus on just waste for this study for several reasons. Firstly this work forms part of a broader project seeking to understand the waste system in Leon, Mexico. More importantly though, is that it was more straightforward to assess the Sustainability of waste behaviors by the students in Mexico, whereas on the food side we deemed to be too complex and individualized in this context.

From that start pointing we made three major changes to the survey instrument for this study. 1) We only asked questions about three knowledge domains having combined social and effectiveness knowledge (as discussed above). 2) Beyond that we sought to shorten the overall instrument to make it easier to deploy both online and for later in-person household studies (not reported here). 3) We needed to make it specific and relevant to Mexico. For example we asked if people “separated” their waste at home. This is not into recycling and non-recycling. Rather the policy in Mexico is generally to get waste separated into “organics” and “inorganics”. The idea is that this keeps potential recyclable material clean in the “inorganics” pile and easy for people later in the waste stream to sort through.

We translated the survey ourselves but not being native Spanish speakers got support on wording from several of our Mexican colleagues. Additionally we piloted the survey with one class of students whose results are not included here due to substantial changes which we made afterwards. We chose to only use 4 Likert options to keep it simpler for the many participants who would have never taken a Likert style survey before. There is mixed evidence for what is the best number of options for a Likert scale and it becomes even murkier when the research is in low-income countries (for a discussion see: http://blogs.worldbank.org/impactevaluations/do-you-agree-or-disagree-how-ask-question). Therefore we omitted the neutral response and reduced the number of options to four from the more standard five or seven. It is important to note that the quantification of the knowledge domains and even the behaviors is not intended to yield precise measurements but instead to focus on broad patterns and relationships between them and between different time periods (e.g. before and after).

Finally it is important to note that this survey shares a weakness common too much of the research in this area in that participants are asked to self-report about knowledge, attitudes and behaviors, rather than these being measured directly. There is evidence that people seem to overestimate their own behaviors (Chao & Lam, 2011) but the effect was not large. Milfont (2009) found that social desirability does not have a strong effect on participant responses as many have feared. While this study makes important advances, we understand its limits and advocate that more future studies measure behaviors directly whenever possible.

B5. Collection of Data

We used Google Forms, a free online program, to administer the survey. See Appendix C for how the survey appeared online. Although youths in Mexico are generally accustomed to using the internet before our classes 2/3 of them had only had online assignments “once and a while”. This was one of the reasons that we sought to keep the survey short and to the point. The survey was the first online assignment the students were given at the beginning of their courses with us. They were given credit for taking the survey but as was made clear to them their responses were anonymized and separate from the course credit. The survey was administered in January and August of 2014 to collect the baseline data before the students had any exposure to Sustainability. In May 2014 at the end of the course the students participating in the intervention were surveyed. Once again taking the survey was a class assignment.

In August 2015 all the students who participated in the intervention were contacted by the email to take a follow-up survey. Unfortunately, it has been our experience that this is not a reliable way to contact students at ENES many of whom do not use email frequently or even at all (but it was our only option). This was a definite barrier to getting the number of responses we might have hoped for and collecting more and better contact information is one of the major methodological areas for improvement in the future. We offered the students who participated a raffle for a $25 Amazon gift card (Amazon had recently begun service in Mexico).

Google Forms automatically creates a spreadsheet of responses as they come in. The spreadsheets were downloaded in CSV format and modified for analysis as described in Appendix D.

B6. Standardization of Results

As mentioned previously in B4 the scales used on the survey are shorter than is typical and not intended to have meaningful quantities. Therefore in order to compare the results of this study in a meaningful way it is necessary that we standardize the outputs. There are many different methods to doing this though most have in common the use of the standard deviation. This gives you a result which tells you how much impact you had relative to the variation between respondents. As was necessary we took slightly different approaches with the regression and the t tests.

For the regressions we produced standardized coefficients based on the methodology proposed by Andrew Gelman (2008). This involves dividing each numeric variable by two times its standard deviation. He argues that this is particularly advantageous when regressions include binary variables (as ours does with for example sex). Additionally each input variable is centered with a mean of zero. These adjustments make the interpretation of regression results much easier-particularly when comparing between items. This procedure has been included in the R-package "arm" under the command *standardize*. In the body of the paper we will report only standardized regression coefficients while the appendix will include both standardized and un-scaled results.

For T tests one common way to standardize your results is to use Cohen's d (Cohen, 1988), which is typically reported as the "effect size" of an intervention. The basic approach involves dividing the mean by the standard deviation. Interpretation is still somewhat subjective but Cohen's suggestion is generally followed: <0.2 "negligible", <0.5 "small", <0.8 "medium", >0.8 "large". For a paired t test such as what we used here there are various proposed approaches to selecting the appropriate standard deviation. We have chosen to use one of the most common which is to divide the difference in the means by the standard deviation of the difference of the means. In order to do that we used the "effsize" package which also produced the accompanying confidence intervals. The body of the paper will report the effect sizes only while the appendix will have both the original t-test results and Cohen's d.

Appendix C: Surveys

C1. Screen Captures of Survey (Spanish)

¿Qué tan familiarizado estás con los siguientes conceptos / términos?
1- "Nunca lo he oído" hasta 4-"Estoy muy familiarizado"

	1	2	3	4
Sostenibilidad (sustentabilidad)	○	○	○	○
Reciclaje	○	○	○	○
Justicia ambiental y social	○	○	○	○
Compostaje	○	○	○	○
Relleno Sanitario	○	○	○	○

¿Cómo calificarías tu capacidad para...
1-"Muy Mal" hasta 4-"Muy Bien"

	1	2	3	4
...hacer decisiones sostenibles de desechos?	○	○	○	○
...influir a los miembros de tu hogar/familia hacer acciones mas sostenibles?	○	○	○	○
...hacer decisiones amigables a medio ambiente de desechos?	○	○	○	○
...influir a tus amigos hacer acciones mas sostenibles?	○	○	○	○
...reducir tu impacto personal al medio ambiente por medio de tus decisiones sobre desechos?	○	○	○	○

¿Cómo calificarías tu acuerdo con las siguientes afirmaciones?
1-"No estoy de acuerdo" hasta 4-"Estoy completamente de acuerdo"

	1	2	3	4
Yo creo que los acciones de un individuo son centrales para lograr la sostenibilidad	○	○	○	○
Yo creo que es fácil reciclar en mi casa.	○	○	○	○
Yo se como hacer compostaje con desechos organics	○	○	○	○
Es importante que todas personas siempre separan sus desechos	○	○	○	○
Yo admiro personas quien se comportan sostenibles.	○	○	○	○
Yo creo que el agua del grifo en León es seguro y sano para beber.	○	○	○	○

¿Cómo calificarías tu acuerdo con las siguientes afirmaciones?

1-"No estoy de acuerdo" hasta 4-"Estoy completamente de acuerdo"

	1	2	3	4
Yo conozco personas que se comprometen a vivir de manera sostenible	○	○	○	○
El plástico ha creado problemas ambientales serios.	○	○	○	○
Salud ambiental es muy importante para justicia social.	○	○	○	○
Me parecerían extraño a otros personas si llevo mi bolsa reusable a la tienda/super para no usar bolsas desechables.	○	○	○	○
Estoy interesado en incorporar la sostenibilidad en mi carrera y futuro profesional.	○	○	○	○
Es la responsabilidad del gobierno para reducir el problema de desechos	○	○	○	○

¿De que frecuencia haces los siguientes acciones?

1-"Nunca" hasta 4-"Siempre"

	1	2	3	4
Yo llevo mi bolsa reusable a la tienda/super para evitar el uso de bolsas desechables.	○	○	○	○
Yo reciclo en mi casa.	○	○	○	○
Yo llevo un botella de agua re-usable a la universidad	○	○	○	○
Yo hago seperacion correcto de mis desechos	○	○	○	○
Yo hago compostaje en mi casa	○	○	○	○

NEXT

C2. English Translation of Survey

a. How familiar are you with the following concepts/terms?
   1-"I have never heard of it" to 4-"I am very familiar with it"
   i. Sustainability
   ii. Recycling
   iii. Social and Environmental Justice
   iv. Composting
   v. Landfill
b. How would you rate your ability to…
   1-"Very bad" to 4-"Very good"
   i. …make sustainable decisions about waste?
   ii. …influence members of my household/family to take more sustainable actions?
   iii. …make environmentally friendly waste decisions?
   iv. …influence your friends to take more sustainable actions?
   v. …reduce your personal impact on the environment by means of your decisions regarding waste?
c. How would you rate your agreement with the following statements?
   1-"I don't agree" to 4-"I agree completely"
   i. I believe that the actions of an individual are central for achieving sustainability.
   ii. I believe that it is easy to recycle in my house.
   iii. I know how to compost with organic waste.
   iv. It is important that everyone always sorts their garbage.
   v. I admire people who behave sustainably.
   vi. I believe that the water from the tap in Leon is safe and healthy to drink
d. How would you rate your agreement with the following statements?
   1-"I don't agree" to 4-"I agree completely"
   i. I know people that are committed to living sustainably.
   ii. Plastic has created serious environmental problems.
   iii. Environmental health is very important for social justice.
   iv. Other people think I am strange if I bring a reusable bag with me to the store so I don't have to use a disposable one.
   v. I am interested in incorporating sustainability in my degree and professional career.
   vi. It is the responsibility of the government to reduce the problem of waste.
e. With what frequency do you do the following actions?
   1-"Never" to 4-"Always"
   i. I bring my reusable bag to the store in order to avoid using a disposable bag.
   ii. I recycle in my house.
   iii. I bring a reusable water bottle with me to campus.
   iv. I correctly sort my garbage.
   v. I compost at home.

C3. Table of Questions with ID as used in Analysis (output from R)

ID	**Question Text**
Q1	X.Qué.tan.familiarizado.estás.con.los.siguientes.conceptos...términos...Sostenibilidad..sustentabilidad..
Q2	X.Qué.tan.familiarizado.estás.con.los.siguientes.conceptos...términos...Reciclaje.
Q3	X.Qué.tan.familiarizado.estás.con.los.siguientes.conceptos...términos...Justicia.ambiental.y.social.
Q4	X.Qué.tan.familiarizado.estás.con.los.siguientes.conceptos...términos...Compostaje.
Q5	X.Qué.tan.familiarizado.estás.con.los.siguientes.conceptos...términos...Relleno.Sanitario.
Q6	X.Cómo.calificarías.tu.capacidad.para.....hacer.decisiones.sostenibles.de.desechos..
Q7	X.Cómo.calificarías.tu.capacidad.para.....influir.a.los.miembros.de.tu.hogar.familia.hacer.acciones.mas.sostenibles..
Q8	X.Cómo.calificarías.tu.capacidad.para.....hacer.decisiones.amigables.a.medio.ambiente.de.desechos..
Q9	X.Cómo.calificarías.tu.capacidad.para.....influir.a.tus.amigos.hacer.acciones.mas.sostenibles..
Q10	X.Cómo.calificarías.tu.capacidad.para.....reducir.tu.impacto.personal.al.medio.ambiente.por.medio.de.tus.decisiones.sobre.desechos..
Q11	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...Yo.creo.que.es.fácil.reciclar.en.mi.casa..
Q12	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...Yo.se.como.hacer.compostaje.con.desechos.organics.
Q13	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...Es.importante.que.todas.personas.siempre.separan.sus.desechos.
Q14	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...Yo.admiro.personas.quien.se.comportan.sostenibles..
Q15	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...Yo.conozco.personas.que.se.comprometen.a.vivir.de.manera.sostenible.
Q16	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...El.plástico.ha.creado.problemas.ambientales.serios..
Q17	X.Cómo.calificarías.tu.acuerdo.con.las.siguientes.afirmaciones...Salud.ambiental.es.muy.importante.para.justicia.social..
Q18	X.De.que.frecuencia.haces.los.siguientes.acciones...Yo.llevo.mi.bolsa.reusable.a.la.tienda.super.para.evitar.el.uso.de.bolsas.desechables..
Q19	X.De.que.frecuencia.haces.los.siguientes.acciones...Yo.reciclo.en.mi.casa..
Q20	X.De.que.frecuencia.haces.los.siguientes.acciones...Yo.llevo.un.botella.de.agua.re.usable.a.la.universidad..
Q21	X.De.que.frecuencia.haces.los.siguientes.acciones...Yo.hago.seperacion.correcto.de.mis.desechos.
Q22	X.De.que.frecuencia.haces.los.siguientes.acciones...Yo.hago.compostaje.en.mi.casa.

C4. Questions on original survey removed before analysis (with justifications)

There were five questions which were included on the survey but which we removed prior to any analysis. The questions were all kept there each time the survey was administered (even if we intended not to include them in the analysis) because we wanted to make sure that the instrument was identical for each person that took it and every time they did so.

Questions Removed

- I believe that the actions of an individual are central for achieving sustainability.
  Initially we had intended to also examine the participants' attitudes about individual agency versus government for making sustainable change but ultimately decided that this was outside the scope of this study.
- It is the responsibility of the government to reduce the problem of waste.
  As per above we decided that this topic area was outside our scope.
- I believe that the water from the tap in Leon is safe and healthy to drink
  We asked this question because we were seeking to collect some initial data to better understand the values and attitudes towards tap water in Leon.
- Other people think I am strange if I bring a reusable bag with me to the store so I don't have to use a disposable one.
  After the initial round of surveys we got feedback that this question was confusing. We discussed the question with Mexican academic colleagues who agreed with that assessment and we therefore decided to drop the question.
- I am interested in incorporating sustainability in my degree and professional career.
  This question was asked just to get a sense of the participants' goals as students in our class and was not intended as part of this study.

Appendix D: Step-by-Step Process of Data Handling and Analysis

D1. Preparing the CSV files:

The following actions were taken with the raw survey data before sharing it.

1. Results from the six surveys were downloaded from Google
2. Some removal of observations because of duplications and significant blank entries
3. Grades and some other demographic information was added
4. Hand coded sex (based on names and memories) and location (standardized and simplified the students' entries
5. Removed irrelevant questions: See Appendix C4
6. Assigned a unique ID to each student and then removed their names to anonymize the data

CSV Files: *Intervention_Presurvey; Intervention_Postsurvey; Intervention_Followupsurvey; Nonintervention_Presurvey*

CSV files are available from the Arizona State University Digital Repository: http://hdl.handle.net/2286/R.C.258

In R

All of the subsequent analysis was done in R. The full R scripts are included in Appendix E which combined with the CSV files should enable anyone to exactly reproduce our results. R software can be downloaded for free here: https://cran.r-project.org/. For a more user friendly experience we recommend the use of RStudio https://www.rstudio.com/. Appendix E is organized in parallel to this so that one could follow along in R.

D2. Creating Necessary Dataframes

In order to analyze the data in R we have to load the CSV files into what R calls dataframes. Most of what we do in this step are various manipulations to get the dataframes set-up for analysis and are not themselves relevant to the research.

We created new variables for each of the four indexes. The mean of the questions included was calculated in order to create this new variable. The questions included in each index are in Table D1. When calculating the means for the indexes questions with no response were omitted, meaning that for some participants their index scores will be composed of fewer questions because they left some answers blank.

Table D1. Questions included in each Index for Analysis	
Declarative	Q1, Q2, Q3, Q4, Q5, Q16, Q17
Procedural	Q6, Q10, Q11, Q12
Subjective	Q7, Q8, Q9, Q13, Q14, Q15
Behavior	Q18, Q19, Q20, Q21, Q22

We decided to treat age as a dummy variable by creating three groups, 18-23 (typical college age), 23-30 (young adult) and 31+ (adult). We thought that it was important to control for age but didn't think it

made sense to do so as a continuous variable (in particular so that the couple participants who were 40+ didn't skew the results). The age ranges we chose were not based on the distribution of results but on what we considered different life stages. In addition grades of students who didn't pass the course were coded as a 0 in the original data so those were all changed to 5 so that it was continuous with the other grades (6 through 10).

D3. Identify and remove outliers

We were very cautious and conservative about removing outliers from the dataset. As mentioned previously some observations were removed because of duplication or large amounts of missing data (ie an entire section was unanswered). To identify outliers in the rest of the data set we calculated Cook's distance which is the most widely used statistical test for measuring the influence of an individual observation (Kim & Storer, 1996). We ran regressions on each of our datasets and then calculated Cook's distance for each observation in each set and created box-plots. The decision to remove outliers and how many was a judgement call. All descriptive statistics and other data reported in this paper has been calculated after the outliers were removed. The outliers have been removed in the process of creating the dataframes in the previous section. If you want to see the original outliers, make sure that you run this section of script right before the section where we removed them. Below Figure D1 is an example boxplot of the results for the 'all student survey' with the outliers we removed identified.

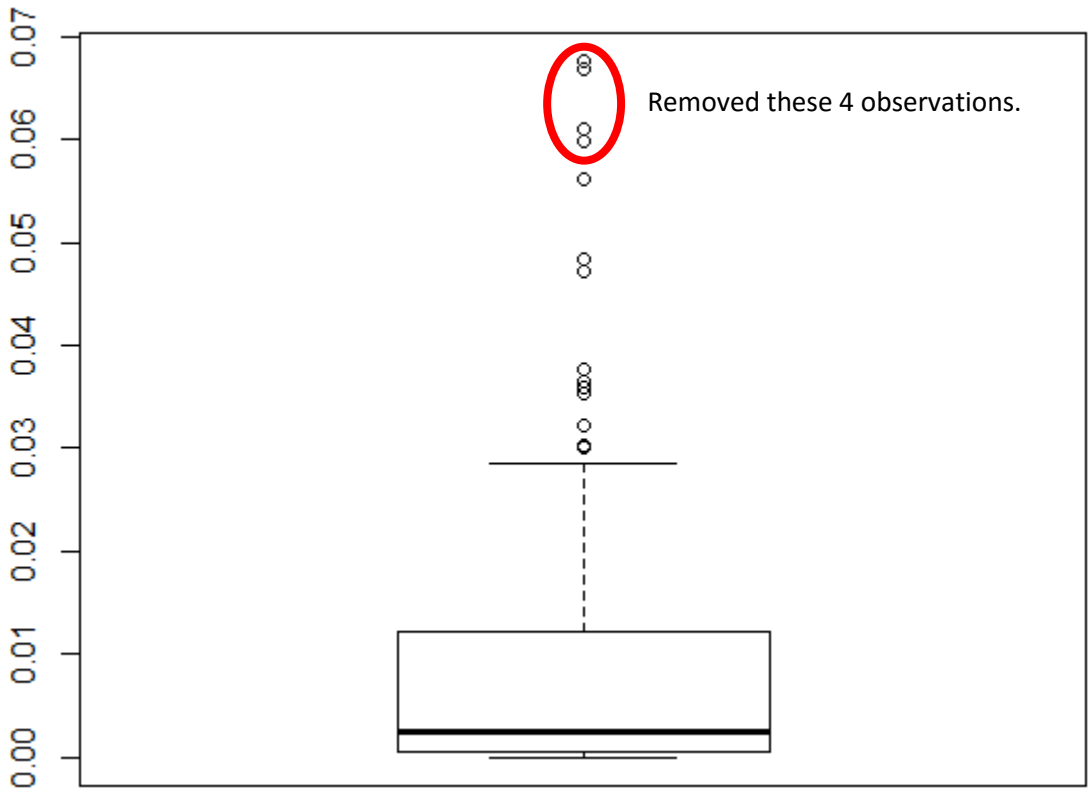


Figure D1. Boxplot of Cook's Distance for observations of 'All Students' regression

D4. Produce Descriptive Statistics of Samples

I couldn't find a package or function in R whose descriptive statistics satisfactory for me. I created my own basic function which can be found in the script Appendix E4. This function takes the percentage which fall into the appropriate category for Sex, Origin, and Degree and the mean of the Age and Grade. Table 2 in the article reports the results of this for the three relevant samples, all of the students, those receiving the intervention and those who followed up one year later. As was noted previously, in preparing the CSV files we hand coded sex and origin. The participants' origin information had been collected but with an open-ended question the results were extremely messy. We cleaned that up and created just four categories for origin: Leon (location of the university), Guanajuato State (not including Leon), Mexico City and Elsewhere.

D5. Compare the samples statistically

We have three samples which we argue are fairly representative of the target population--students at ENES. This population is probably broadly similar to other university student populations with specific differences with UNAM (its parent campus) because there the vast majority of students are from Mexico City and differs from other private and public universities in Mexico because UNAM/ENES is by far the most selective. Although the sampling methodology was not random we contend that our samples are sufficiently representative of the student body. The broadest sample 'All Students' captures a cross section of students from virtually all the degree programs, reflecting the composition of the student body. Our other two samples are sequential sub-samples of this.

We used the Kolmogorov-Smirnov (KS) test to evaluate whether the sub-samples were drawn from the same population as the students who were not part of those sub-samples. In particular we are concerned with whether the sample has significantly different attitudes, knowledge or behaviors with regards to Sustainability. First we compared the students who participated in the intervention with those who did not in Table D2.

Table D2. KS Test comparing Intervention and Nonintervention Samples

	Intervention (mean)	Nonintervention (mean)	KS Statistic	KS P-Value
Declarative	3.09	2.99	0.12	0.84
Procedural	2.82	2.84	0.10	0.96
Subjective	3.17	3.11	0.09	0.99
Behavior	2.53	2.50	0.08	1

Visually, one can see that the mean scores of students who participated in the intervention and those that didn't were very close together and the KS test confirms that statistically these two samples are representative of the same population.

For the next subsample, slightly more than a third of the participants in the intervention responded to the follow-up survey. Table D3 below shows that those that followed up and those that didn't are statistically similar though there is some concern about the difference in grades, which while not quite significant is apparent and perhaps not surprising.

Table D3. KS Test comparing Preintervention sample with Followup

	Followups (mean)	Non-Followups (mean)	KS Statistic	KS P-Value
Declarative	3.19	3.03	0.23	0.26
Procedural	2.83	2.81	0.15	0.78

Subjective	3.16	3.17	0.12	0.94
Behavior	2.58	2.51	0.06	1
Final Grade	8.07	7.61	0.29	0.10

D6. Assessing the indices with Cronbach's Alpha

This research investigated the relationship of domains of knowledge and behavior. We assessed this creating indices for each knowledge domain and the behaviors composed of a set of questions. The expectation is that the answers to these questions should be correlated with each other. Cronbach's Alpha reliability coefficient is a widely utilized method for measuring the correlation between items composing an index (Bland & Altman, 1997). The scores range from 0 to 1 with higher scores indicating a greater correlation between the index's items. Table D4 reports the Cronbach's Alpha coefficient in standardized form for the three samples.

Table D4. Standardized Cronbach's Alpha for indices in each sample

	All Students	Post-Intervention	Follow-up
Declarative	0.65	0.76	0.83
Procedural	0.60	0.45	0.78
Subjective	0.70	0.64	0.83
Behavior	0.46	0.70	0.78

The way that Cronbach's Alpha changes for each of the indices in each of the studied groups is very interesting but we are unsure of how to interpret this result. I am for example not surprised that the behavior index was relatively uncorrelated initially as we asked about a diverse set of waste related actions but that after the intervention the students saw them more as connected (e.g. using a re-usable water bottle as a waste reduction strategy). Another potential interpretation from this table is that after the intervention respondents are more inconsistent with their responses but then over the long term they settle into more consistency thus the high coefficients in the follow-up survey.

D7. Multicollinearity among independent variables in the regressions

In order to check whether the independent variables in our regressions were correlated we calculated the Variance Influence Factor (VIF) for all the regressions. VIF is widely used to assess multicollinearity and typically a score above 4 is considered to be a concern while 10 indicates multicollinearity that probably should be corrected (O'Brien, 2007). Below Table D5. contains the VIF for all of the regressions with the variables of concern highlighted. As can be seen the only concern with multicollinearity is in the follow-up sample with procedural and declarative knowledge. Extra caution is therefore taken with interpreting the results from that particular regression.

Table D5. Variance Influence Factors for all Regressions

	All Students	Pre Intervention	Post Intervention	Followup	Differences Pre/Post	Differences Pre/Followup
Sex	1.301	1.395	1.292	2.276	1.360	2.270
AgeCategory	1.379	1.824	1.394	2.464	1.438	2.586
Origin	1.633	1.681	1.507	3.015	1.629	3.043
Degree	1.931	2.126	2.172	3.487	1.922	2.329
Final.Grade	1.486	1.770	1.407	1.894	1.497	2.326

Declarative	1.559	1.676	1.529	5.348	1.745	3.361
Procedural	1.996	2.316	1.626	9.471	1.913	4.362
Subjective	1.564	2.070	1.781	3.835	2.226	1.977

D8. Control Variables

For our various regression analysis we decided to include the same set of control variables: sex, age, origin, degree, and final grade. The following is a brief justification for why we believed these to be important to include:

- Sex: Other studies have found that women are more likely than men to report pro-environmental attitudes and behaviors (Christensen & Knezek, 2015; Gossard & York, 2003; Sahin, Ertepinar, & Teksoz, 2012). Indeed, in the initial survey did find that result.
- Age: As previously mentioned age was converted into three categories. Basically we view age as being relevant in terms of one's living circumstances and life experiences. The college age group has probably never lived on their own or is just doing it for the first time (far less common in Mexico than the US) and therefore may never have managed household waste before. The young adult group is more likely to have lived on their own and certainly held a job, though many still probably lived with families. The oldest group is by far the most likely to have managed their own household waste.
- Origin: The participant's origin is of importance because different cities in Mexico manage their waste very differently. For example, parts of Mexico City have separate garbage, recycling and organic pick-ups whereas Leon only has a garbage pick-up (and some places may have no pick-up).
- Degree: The background of the students in the different degree programs are somewhat different. More importantly students in each degree program take 100% classes within those programs. Potentially the students from the different programs could be statistically different. In fact it appears that the students from the Intercultural Development and Management degree program are, though since only one of them participates in the intervention, this does not play much of a role in this study.

D9. Evaluating Hypothesis 1

H_1: Higher levels of subjective and procedural knowledge correlates with more sustainable behavior while higher levels of declarative knowledge does not.

To assess our first hypothesis we studied the full sample of students before they had any Sustainability related intervention (ie during the first weeks of the semester). In table D6. are the full regression results for every sample and the differences between samples, while table D7. is the standardized results.

D6. Summary of all regression results in study

	Dependent variable:					
	Behavior					
	All Students	Pre Intervention	Post Intervention	Followup	Differences Pre/Post	Differences Pre/Followup
	(1)	(2)	(3)	(4)	(5)	(6)
Sexf	0.165**	0.209*	0.036	0.049	-0.050	-0.251
	(0.080)	(0.106)	(0.135)	(0.212)	(0.140)	(0.224)
AgeCategory(23,30]	0.165	0.149	0.052	0.158	-0.116	-0.031
	(0.123)	(0.179)	(0.259)	(0.322)	(0.258)	(0.334)
AgeCategory(30,100]	0.169	0.181	0.054	-0.028	-0.170	-0.321
	(0.236)	(0.257)	(0.392)	(0.303)	(0.387)	(0.334)
OriginLeon	0.095	0.158	-0.047	0.106	-0.024	0.050
	(0.100)	(0.133)	(0.178)	(0.198)	(0.182)	(0.229)
OriginGTO	0.190	0.227	-0.144	0.290	-0.131	0.304
	(0.116)	(0.147)	(0.190)	(0.237)	(0.196)	(0.272)
OriginDF	0.207	0.266	0.173	0.097	0.192	0.019
	(0.135)	(0.175)	(0.227)	(0.269)	(0.240)	(0.270)
DegreeDesarrollo y Gestión Interculturales	0.257*	0.924*				
	(0.153)	(0.467)				
DegreeEconomia Industrial	-0.013					
	(0.138)					
DegreeFisioterapía	0.069	0.152	-0.101	-0.075	-0.147	0.110
	(0.117)	(0.172)	(0.226)	(0.355)	(0.220)	(0.348)
DegreeOdontología	0.107	0.183	-0.100	-0.150	-0.338	-0.224
	(0.125)	(0.164)	(0.231)	(0.318)	(0.224)	(0.329)

Final.Grade	0.010	0.014	-0.014	-0.099	-0.043	-0.110
	(0.033)	(0.045)	(0.056)	(0.077)	(0.060)	(0.091)
Declarative	0.122	0.095	0.083	0.491	0.155	0.058
	(0.101)	(0.131)	(0.200)	(0.303)	(0.144)	(0.258)
Procedural	0.304***	0.322**	0.439**	-0.025	0.107	0.174
	(0.098)	(0.137)	(0.195)	(0.314)	(0.151)	(0.220)
Subjective	0.374***	0.296**	0.668***	0.724***	0.498**	0.497*
	(0.101)	(0.147)	(0.238)	(0.235)	(0.194)	(0.235)
Constant	-0.269	-0.170	-1.044	-0.408	0.659	1.168
	(0.421)	(0.527)	(0.872)	(0.740)	(0.471)	(0.678)
Observations	119	82	75	28	72	28
R^2	0.466	0.476	0.383	0.827	0.362	0.648
Adjusted R^2	0.395	0.376	0.264	0.688	0.233	0.366
Residual Std. Error	0.379 (df = 104)	0.404 (df = 68)	0.508 (df = 62)	0.335 (df = 15)	0.500 (df = 59)	0.356 (df = 15)
F Statistic	6.493*** (df = 14; 104)	4.754*** (df = 13; 68)	3.212*** (df = 12; 62)	5.959*** (df = 12; 15)	2.794*** (df = 12; 59)	2.302* (df = 12; 15)
Note:						*p<0.1; **p<0.05; ***p<0.01

D7. Summary of standardized regression results in study

	Dependent variable:					
	Behavior					
	All Students	Pre Intervention	Post Intervention	Followup	Differences Pre/Post	Differences Pre/Followup
	(1)	(2)	(3)	(4)	(5)	(6)
c.Sex	0.165**	0.209*	0.036	0.049	-0.050	-0.251
	(0.080)	(0.106)	(0.135)	(0.212)	(0.140)	(0.224)
AgeCategory(23,30]	0.165	0.149	0.052	0.158	-0.116	-0.031
	(0.123)	(0.179)	(0.259)	(0.322)	(0.258)	(0.334)

AgeCategory(30,100]	0.169	0.181	0.054	-0.028	-0.170	-0.321
	(0.236)	(0.257)	(0.392)	(0.303)	(0.387)	(0.334)
OriginLeon	0.095	0.158	-0.047	0.106	-0.024	0.050
	(0.100)	(0.133)	(0.178)	(0.198)	(0.182)	(0.229)
OriginGTO	0.190	0.227	-0.144	0.290	-0.131	0.304
	(0.116)	(0.147)	(0.190)	(0.237)	(0.196)	(0.272)
OriginDF	0.207	0.266	0.173	0.097	0.192	0.019
	(0.135)	(0.175)	(0.227)	(0.269)	(0.240)	(0.270)
DegreeDesarrollo y Gestión Interculturales	0.257*	0.924*				
	(0.153)	(0.467)				
DegreeEconomia Industrial	-0.013					
	(0.138)					
DegreeFisioterapía	0.069	0.152	-0.101	-0.075	-0.147	0.110
	(0.117)	(0.172)	(0.226)	(0.355)	(0.220)	(0.348)
DegreeOdontología	0.107	0.183	-0.100	-0.150	-0.338	-0.224
	(0.125)	(0.164)	(0.231)	(0.318)	(0.224)	(0.329)
z.Final.Grade	0.026	0.038	-0.034	-0.229	-0.103	-0.253
	(0.085)	(0.119)	(0.140)	(0.178)	(0.145)	(0.209)
z.Declarative	0.106	0.084	0.061	0.483	0.168	0.057
	(0.087)	(0.116)	(0.146)	(0.298)	(0.157)	(0.251)
z.Procedural	0.305***	0.320**	0.339**	-0.031	0.117	0.227
	(0.099)	(0.137)	(0.151)	(0.397)	(0.164)	(0.286)
z.Subjective	0.324***	0.261**	0.443***	0.777***	0.455**	0.407*
	(0.087)	(0.129)	(0.158)	(0.253)	(0.177)	(0.193)
Constant	2.315***	2.197***	3.066***	2.749***	0.636**	0.160
	(0.134)	(0.190)	(0.255)	(0.360)	(0.254)	(0.388)

Observations	119	82	75	28	72	28
R^2	0.466	0.476	0.383	0.827	0.362	0.648
Adjusted R^2	0.395	0.376	0.264	0.688	0.233	0.366
Residual Std. Error	0.379 (df = 104)	0.404 (df = 68)	0.508 (df = 62)	0.335 (df = 15)	0.500 (df = 59)	0.356 (df = 15)
F Statistic	6.493^{***} (df = 14; 104)	4.754^{***} (df = 13; 68)	3.212^{***} (df = 12; 62)	5.959^{***} (df = 12; 15)	2.794^{***} (df = 12; 59)	2.302^{*} (df = 12; 15)
Note:						*p<0.1; **p<0.05; ***p<0.01

D10. Evaluating Hypothesis 2

H_2: Participating students will show an increase in their procedural and subjective knowledge as well as sustainable behaviors after the educational program.

To be able to assess this hypothesis we did a paired t-test comparing the pre answers with the post answers. To be able to do this we had to match the pre database with the post database as not all the students completed one or the other. There were 72 paired observations which we used (also had removed outliers as previously discussed). It took some manipulation to get the results in a usable format. Table D8. contains the results from the t-test and the Cohen's D effect size calculations for all the questions (Cohen's D is not calculated for all the questions due to some blank responses).

Table D8. T-Test Results and Cohen's D for Pre vs. Post Changes in Survey for All Questions

Names	estimate	p.value	Adjusted.P.Values	conf.int_min	conf.int_max	Tstatistic	Degrees Freedom	Cohens D	CD.conf.int_min	CD.conf.int_max
Q1	0.887	0	0	0.697	1.077	9.315	70			
Q2	0.222	0.002	0.010	0.084	0.360	3.214	71	0.379	0.044	0.714
Q3	0.486	0.00004	0.0002	0.267	0.706	4.415	71	0.520	0.183	0.858
Q4	0.806	0.00000	0.00000	0.522	1.089	5.667	71	0.668	0.327	1.009
Q5	0.819	0	0.00000	0.575	1.064	6.692	71	0.789	0.444	1.133
Q6	0.667	0	0	0.482	0.852	7.186	71	0.847	0.500	1.193
Q7	0.319	0.001	0.005	0.135	0.504	3.457	71	0.407	0.072	0.743
Q8	0.486	0.00000	0.00000	0.319	0.653	5.794	71	0.683	0.341	1.024
Q9	0.443	0.00002	0.0002	0.250	0.636	4.574	69			
Q10	0.361	0.0002	0.001	0.179	0.543	3.955	71	0.466	0.130	0.802
Q11	0.181	0.027	0.104	0.021	0.340	2.259	71	0.266	-0.067	0.600
Q12	1.211	0	0	0.931	1.491	8.633	70			
Q13	0.194	0.012	0.050	0.044	0.345	2.569	71	0.303	-0.031	0.637
Q14	0.268	0.00003	0.0002	0.148	0.387	4.457	70			
Q15	0.653	0.00001	0.0001	0.382	0.924	4.806	71	0.566	0.228	0.905
Q16	0.375	0.0001	0.0005	0.197	0.553	4.194	71	0.494	0.157	0.831
Q17	0.282	0.003	0.015	0.098	0.466	3.051	70			
Q18	0.569	0.00000	0.00005	0.340	0.799	4.950	71	0.583	0.245	0.922
Q19	0.278	0.003	0.012	0.100	0.455	3.123	71	0.368	0.033	0.703
Q20	0.125	0.118	0.435	-0.032	0.282	1.583	71	0.187	-0.146	0.519
Q21	0.352	0.007	0.030	0.099	0.605	2.772	70			
Q22	0.722	0.00000	0.00003	0.443	1.002	5.152	71	0.607	0.268	0.947

Figure D2. is a graphical result showing the estimated difference after the intervention in the survey answers with 95% confidence intervals.

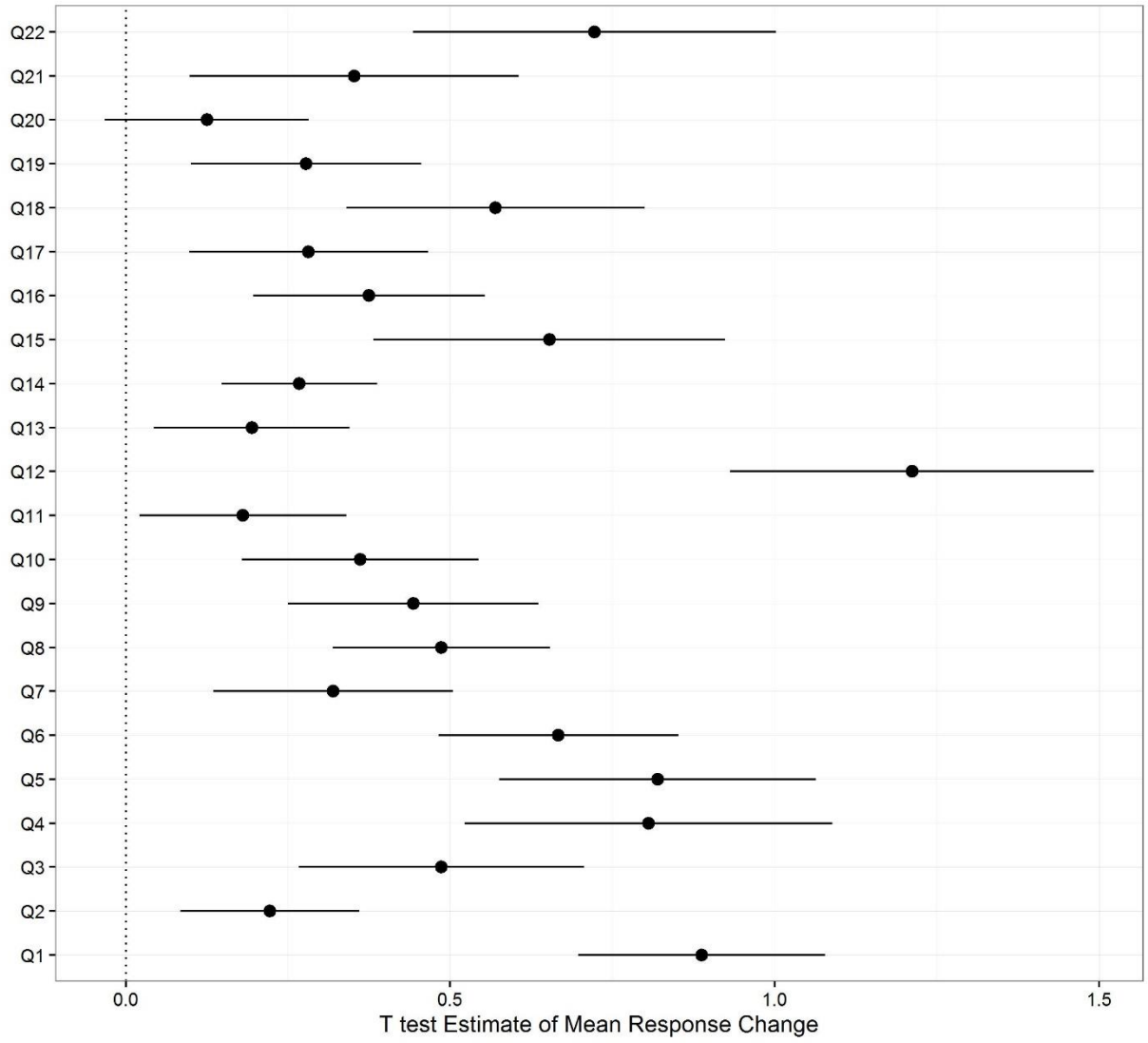


Figure D2. Estimate of mean difference for all questions with confidence intervals

Table D9. T-Test Results and Cohen's D for Pre vs. Post Changes in Survey for Indices

Names	estimate	p.value	Adjusted. P.Values	conf.int _min	conf.int _max	Tstatistic	DF	Cohe nsD	CD.conf.i nt_min	CD.conf.i nt_max
Declarative	0.553	0	0	0.425	0.680	8.645	71	1.019	0.666	1.372
Procedural	0.601	0	0	0.474	0.729	9.381	71	1.106	0.749	1.462
Subjective	0.394	0	0	0.287	0.502	7.321	71	0.863	0.516	1.210
Behavior	0.411	0.00000	0.00000	0.277	0.545	6.114	71	0.721	0.378	1.063

D11. Adjusting P-values for multiple comparisons

When you are making multiple comparisons your risk for Type II error (false positive) goes up for each additional comparison you make. Schochet (2008) lays out guidelines for dealing with this issue in evaluations of educational interventions and suggest two remedies. One is to use indexes instead of individual questions to reduce the number of comparisons you are doing. The comparison of the indexes is the main point of our analysis and what we focused on in the article. The other solutions involves adjusting your p-values to account for the number of comparisons. There are various methodologies have been developed for this adjusting which involve different assumptions and trade-offs about your data. We chose to adjust the p-values with the method developed by Benjamini and Yekutieli (2016). This has been included in all of the previous tables.

D12. Evaluating Hypothesis 3

H_3: Changes in procedural and subjective knowledge will predict changes in sustainable behavior while changes in declarative knowledge will not.

In order to evaluate this hypothesis we created a new dataframe using the matching pre and post dataframes we had created for hypothesis 2. The pre scores were subtracted from the post scores for each participant yielding a measure of the change that occurred due to the intervention. A regression was run on this data to look for relationships with the dependent variable of changing behavior. See the results in tables D6 and D7 above.

D13. Evaluating Hypothesis 4

H_4: After one year participating students will still have increased knowledge (in all domains) and sustainable behaviors relative to before the education program.

Unfortunately, we were only able to collect responses one year later from about one third of the participants, which after processing and removing outliers left us with 28 observations to compare to the pre-intervention results. After adjusting for multiple comparisons there are only three questions for which we can be claim that there was a long term change, see Table D10.

Table D10. T-Test Results and Cohen's D for Pre vs. Follow-up Changes for All Questions

Names	estimate	p.value	Adjusted. P.Values	conf.int _min	conf.int _max	Tstatistic	Degrees Freedom	CohensD	CD.conf.int _min	CD.conf.int_ max
Q1	0.571	0.0004	0.029	0.284	0.859	4.076	27	0.770	0.204	1.337
Q2	0.107	0.264	1	-0.086	0.300	1.140	27	0.215	-0.332	0.763
Q3	0.036	0.851	1	-0.352	0.423	0.189	27	0.036	-0.510	0.581
Q4	0.536	0.007	0.132	0.163	0.909	2.948	27	0.557	0.001	1.114
Q5	0.714	0.001	0.036	0.321	1.107	3.731	27	0.705	0.142	1.268
Q6	0.107	0.523	1	-0.232	0.446	0.648	27	0.122	-0.424	0.669
Q7	-0.214	0.184	1	-0.537	0.109	-1.362	27	-0.257	-0.805	0.291
Q8	0.286	0.073	0.986	-0.028	0.600	1.867	27	0.353	-0.197	0.903
Q9	-0.214	0.110	1	-0.480	0.052	-1.652	27	-0.312	-0.861	0.237
Q10	0.036	0.823	1	-0.289	0.361	0.225	27	0.043	-0.503	0.588
Q11	-0.071	0.626	1	-0.369	0.226	-0.493	27	-0.093	-0.639	0.453
Q12	0.778	0.002	0.046	0.321	1.235	3.500	26			
Q13	-0.071	0.490	1	-0.281	0.138	-0.701	27	-0.132	-0.679	0.414
Q14	0.071	0.424	1	-0.109	0.252	0.812	27	0.153	-0.393	0.700
Q15	0.250	0.148	1	-0.094	0.594	1.491	27	0.282	-0.267	0.830
Q16	0.071	0.646	1	-0.244	0.387	0.465	27	0.088	-0.458	0.634
Q17	-0.107	0.501	1	-0.430	0.215	-0.682	27	-0.129	-0.675	0.417
Q18	0.357	0.057	0.929	-0.012	0.726	1.987	27	0.375	-0.175	0.926
Q19	0.071	0.646	1	-0.244	0.387	0.465	27	0.088	-0.458	0.634
Q20	0.148	0.381	1	-0.194	0.490	0.891	26			
Q21	0.250	0.183	1	-0.125	0.625	1.368	27	0.259	-0.289	0.807

Q22	0.250	0.129	1	-0.077	0.577	1.567	27	0.296	-0.253	0.845

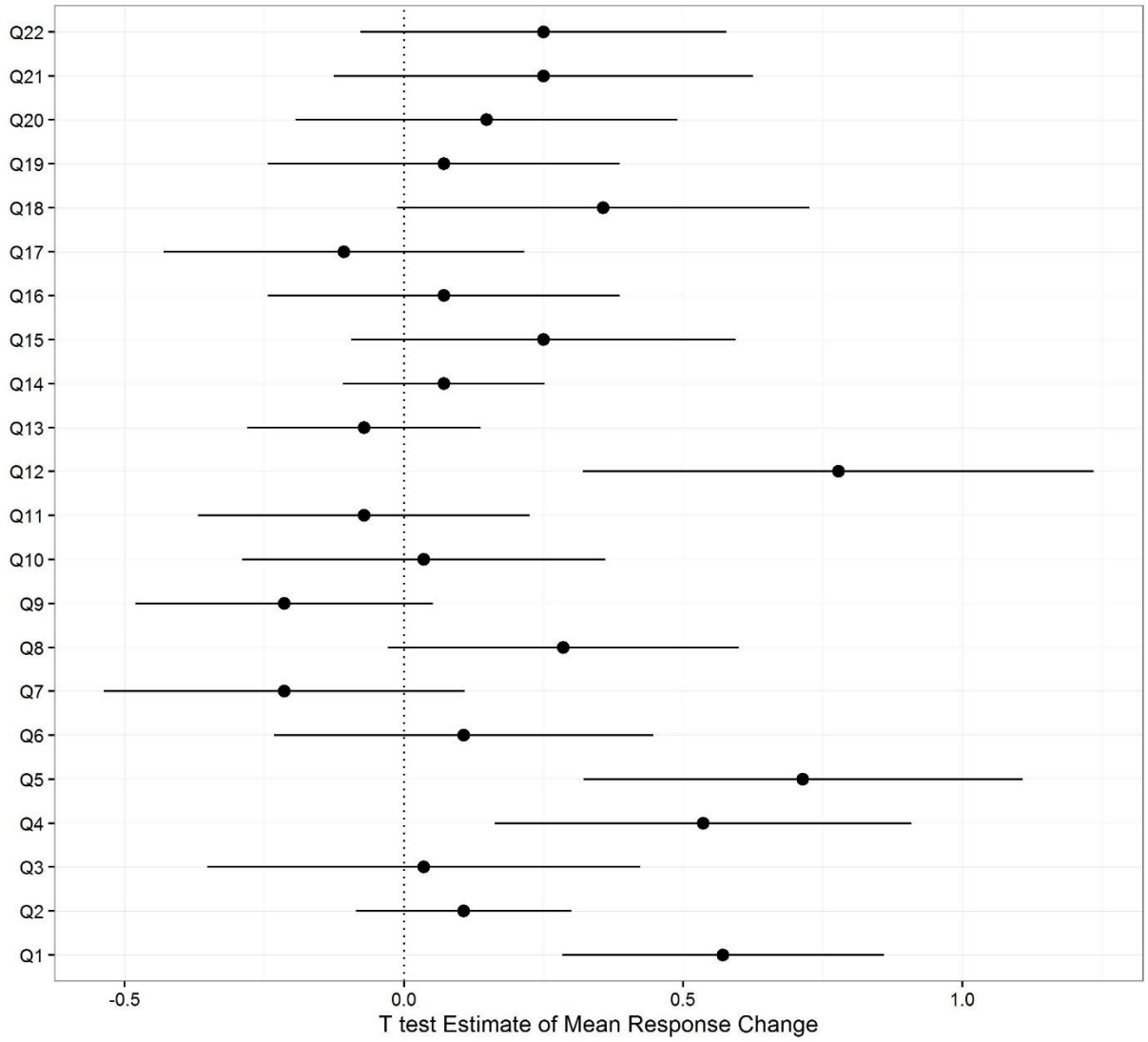


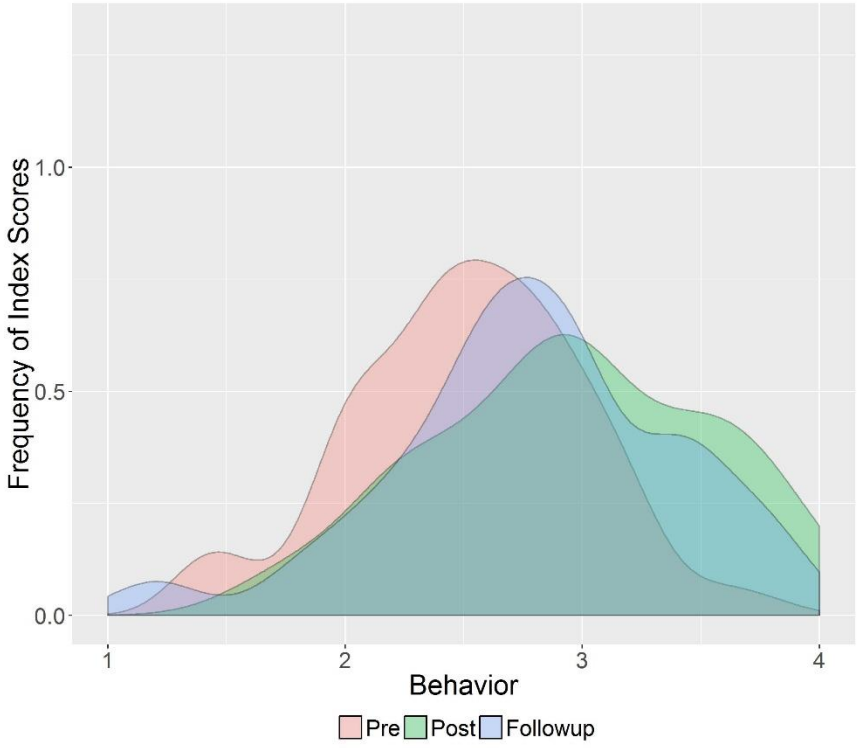
Figure D3. Estimate of mean difference for all questions with confidence intervals Pre vs. Follow-up

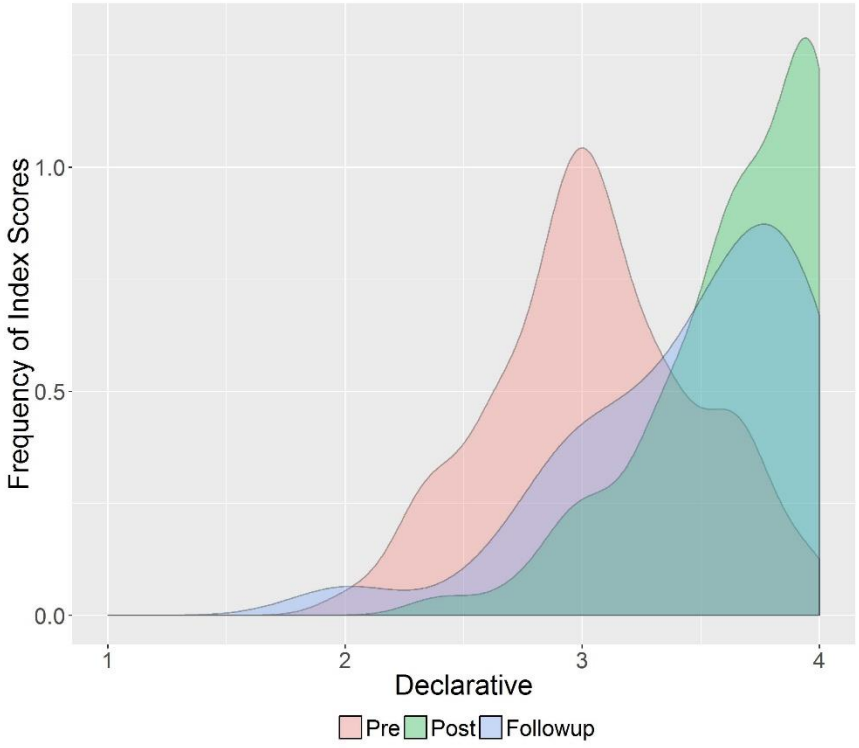
We can also look at the indices which once adjusting the p-values we can see a sure change in Declarative knowledge but less confidence that there was a change in behavior.

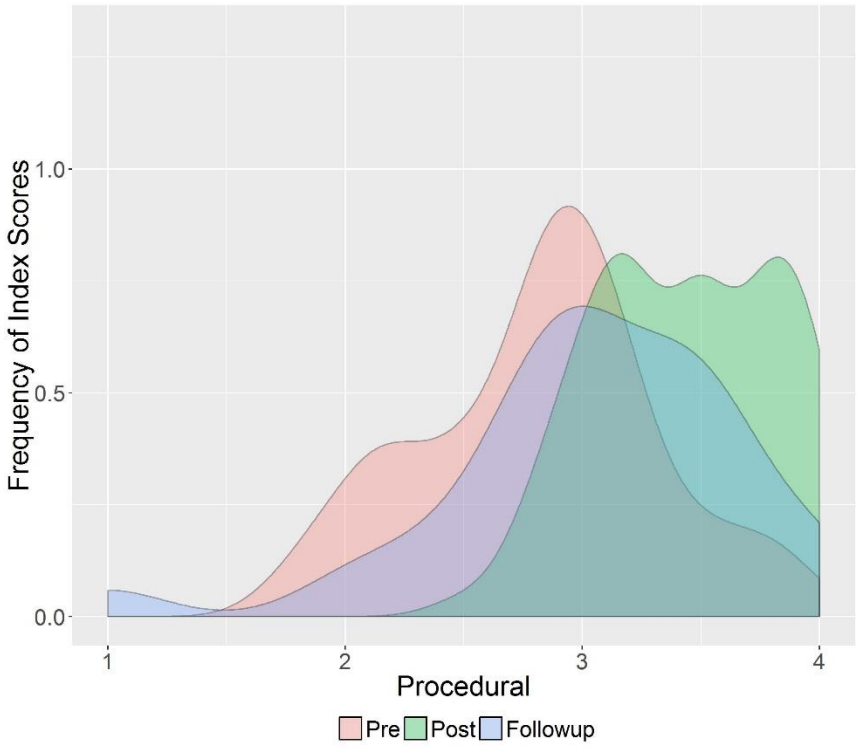
Table D11. T-Test Results and Cohen's D for Pre vs. Follow-up Changes in Indices

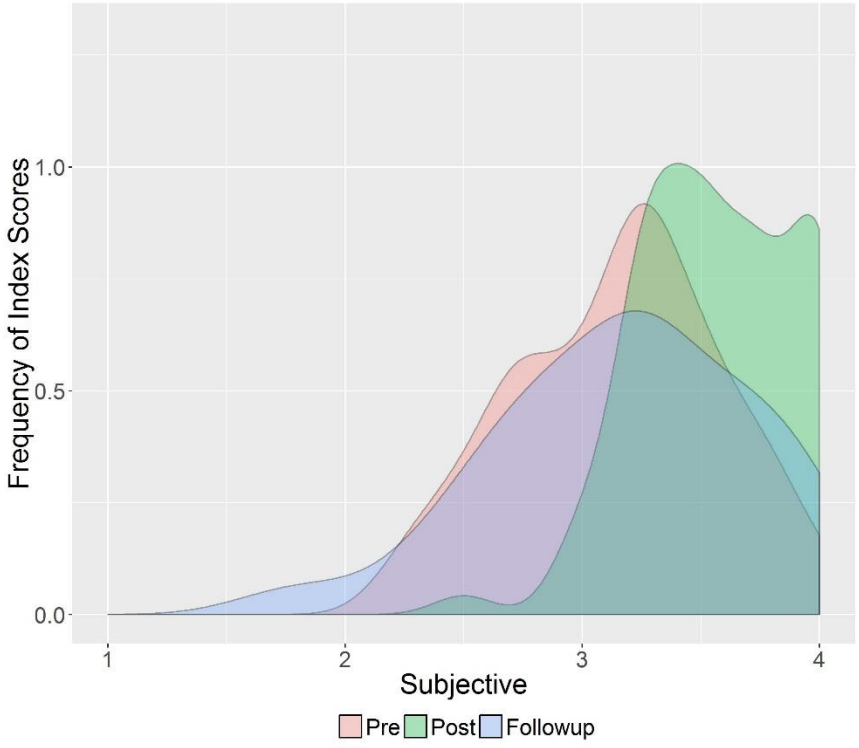
Names	estimate	p.value	Adjusted. P.Values	conf.int _min	conf.int _max	Tstatistic	Degrees Freedom	Cohens D	CD.conf. int_min	CD.conf. int_max
Declarative	0.282	0.005	0.041	0.093	0.471	3.066	27	0.579	0.022	1.137
Procedural	0.214	0.093	0.258	-0.038	0.467	1.742	27	0.329	-0.220	0.879
Subjective	0.011	0.891	1	-0.148	0.169	0.138	27	0.026	-0.520	0.572
Behavior	0.218	0.016	0.066	0.044	0.391	2.577	27	0.487	-0.067	1.041

D14. Density Plot Histograms of the Indices Figure D4.






Frequency of Index Scores
1.0
0.5
0.0
1
2
3
4
Procedural
Pre
Post
Followup


Frequency of Index Scores
1.0
0.5
0.0
1
2
3
4
Subjective
Pre
Post
Followup

D15. Investigation into the Sex Differences

The results were suggestive of a difference between the sexes, at least on the all student sample. This difference can be seen in Figure D5.

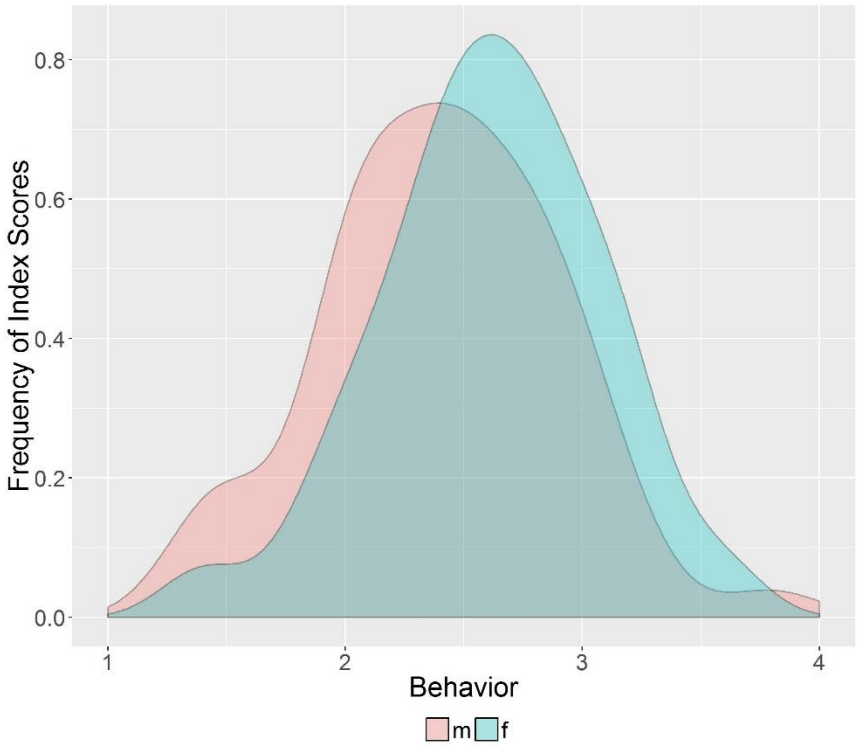


Figure D5. Density Plot Histogram of the Behavior in the All Student Sample Split by Sex

In order to compare the differences between the sexes statistically we split the databases into male and female ones and then ran a t-test comparing the responses for behavior on each of the three samples. The results of which can be seen in the table D12. below. This finds only a significant difference in the 'All Student' sample.

Table D12. Unpaired T-test of Females vs. Males for Each Sample

Surveys	t.statistic	parameter	p.value	FemaleMean	MaleMean	DiffofMeans	confint_min	confint_max
All	2.291	106.083	0.024	2.612	2.408	0.204	0.028	0.381
Post	-0.803	62.406	0.425	2.900	3.013	-0.113	-0.394	0.168
Followup	1.033	11.884	0.322	2.870	2.600	0.270	-0.300	0.840

These data show a statistically significant difference between the female and male respondents initially but that difference is no longer there in the later surveys. We examined this a different way using the Kolmogorov-Smirnov (KS) to ask whether the males and females were drawn from different populations. In Table D13. one can see that for none of the indices can we conclude that males and females are different populations even if the means are statistically different as we found with the t-test.

Table D13. KS Test comparing Females vs Males for All student Sample

	Female (mean)	Male (mean)	KS Statistic	KS P-Value
Declarative	3.07	3.04	0.10	0.95
Procedural	2.87	2.76	0.12	0.83
Subjective	3.21	3.08	0.18	0.29
Behavior	2.61	2.41	0.19	0.25

We ran a regression on the male and female samples for the all student survey data and report the results in the table D14 and the standardized results in D15. The most interesting result is the difference between the effect of subjective and procedural knowledge for the sexes. The effect is still there in the post survey but dividing that already smaller sample dramatically reduces the power of that result.

Table D14. Regressions of All Student Survey Separated by Sex

	Dependent variable:	
	Behavior	
	Female	Male
	(1)	(2)
AgeCategory(23,30]	0.236	0.103
	(0.206)	(0.154)
AgeCategory(30,100]	0.372	0.095
	(0.392)	(0.306)
OriginLeon	0.176	-0.010
	(0.126)	(0.159)
OriginGTO	0.079	0.269
	(0.138)	(0.197)
OriginDF	0.586**	-0.045
	(0.244)	(0.175)
DegreeDesarrollo y Gestión Interculturales	0.426*	0.062
	(0.220)	(0.213)

DegreeEconomia Industrial	-0.070	-0.227
	(0.171)	(0.235)
DegreeFisioterapía	0.191	-0.145
	(0.145)	(0.191)
DegreeOdontología	0.290*	-0.217
	(0.156)	(0.194)
Final.Grade	-0.013	0.015
	(0.045)	(0.057)
Declarative	0.123	0.141
	(0.121)	(0.172)
Procedural	0.490***	0.048
	(0.119)	(0.174)
Subjective	0.110	0.792***
	(0.128)	(0.159)
Constant	0.267	-0.630
	(0.547)	(0.678)
Observations	67	52
R^2	0.523	0.605
Adjusted R^2	0.406	0.470
Residual Std. Error	0.358 (df = 53)	0.361 (df = 38)
F Statistic	4.463*** (df = 13; 53)	4.476*** (df = 13; 38)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table D15. Standardized Regressions of All Student Survey Separated by Sex

	Dependent variable:	
	Behavior	
	Female	Male
	(1)	(2)
AgeCategory(23,30]	0.236	0.103
	(0.206)	(0.154)
AgeCategory(30,100]	0.372	0.095
	(0.392)	(0.306)
OriginLeon	0.176	-0.010
	(0.126)	(0.159)
OriginGTO	0.079	0.269
	(0.138)	(0.197)
OriginDF	0.586**	-0.045
	(0.244)	(0.175)
DegreeDesarrollo y Gestión Interculturales	0.426*	0.062
	(0.220)	(0.213)
DegreeEconomia Industrial	-0.070	-0.227
	(0.171)	(0.235)
DegreeFisioterapía	0.191	-0.145
	(0.145)	(0.191)
DegreeOdontología	0.290*	-0.217
	(0.156)	(0.194)
z.Final.Grade	-0.030	0.043
	(0.102)	(0.160)

z.Declarative	0.108	0.121
	(0.106)	(0.148)
z.Procedural	0.506***	0.046
	(0.123)	(0.166)
z.Subjective	0.098	0.648***
	(0.114)	(0.131)
Constant	2.292***	2.487***
	(0.160)	(0.219)
Observations	67	52
R^2	0.523	0.605
Adjusted R^2	0.406	0.470
Residual Std. Error	0.358 (df = 53)	0.361 (df = 38)
F Statistic	4.463*** (df = 13; 53)	4.476*** (df = 13; 38)
Note:		*p<0.1; **p<0.05; ***p<0.01

Appendix E: R Script to Reproduce Analysis

E1. Loading the CSV files into R

```
#Packages Used (in the end some of these are probably not necessary
but we are not sure which ones)
library(dplyr)
library(tidyr)
library(ggplot2)
library(psych)
library(splitstackshape)
library(car)
library(stargazer)
library(hexbin)
library(Hmisc)
library(arm)
library(effsize)

#Set your working directory to location of CSV files and all R-scripts
setwd("……")

#We load all the scripts from a folder named Rscripts, create this
folder in your working directory or change the file name saves

#Load CSV files into R
PreIntervention <- read.csv("Intervention_Presurvey.csv")
PostIntervention <- read.csv("Intervention_Postsurvey.csv")
FollowIntervention <- read.csv("Intervention_Followupsurvey.csv")
NonIntervention <- read.csv("Nonintervention_Presurvey.csv")
```

E2. Creating the necessary dataframes

```
#Record Questions in Dataframe
Questions <- data.frame(paste("Q",1:22,
sep=""),names(PreIntervention[,8:29]))
colnames(Questions) <- c("ID","Question Text")

#Change column names for questions to Q1-Q22
#We are doing this because the full questions are unwiedingly long
colnames(PreIntervention)[8:29] <- paste("Q", 1:22, sep="")
colnames(PostIntervention)[8:29] <- paste("Q", 1:22, sep="")
colnames(FollowIntervention)[8:29] <- paste("Q", 1:22, sep="")
colnames(NonIntervention)[8:29] <- paste("Q", 1:22, sep="")

#Simplifying and Putting into English Demographic Column Names
colnames(PreIntervention)[3:6] <- c("Age","Origin","Degree","Year")
colnames(PostIntervention)[3:6] <- c("Age","Origin","Degree","Year")
colnames(FollowIntervention)[3:6] <- c("Age","Origin","Degree","Year")
colnames(NonIntervention)[3:6] <- c("Age","Origin","Degree","Year")

#Creating a vector of names we can use for graphs
```

```
xaxis_names <-c(paste("Q",1:22, sep=""),"D","P","S","B")

#Next we are going to create composite variables for each "Knowledge
Domain"
#These vectors identify which questions go in each domain
DeclarativeVec <- c("Q1","Q2","Q3","Q4","Q5","Q16","Q17")
ProceduralVec <- c("Q6","Q10","Q11","Q12")
SubjectiveVec <- c("Q7","Q8","Q9","Q13","Q14","Q15")
BehaviorVec <- c("Q18","Q19","Q20","Q21","Q22")

#Print Table
Indexes <- rbind(paste(DeclarativeVec,collapse=",
"),paste(ProceduralVec,collapse=", "),paste(SubjectiveVec,collapse=",
"),paste(BehaviorVec,collapse=", "))
row.names(Indexes) <-
c("Declarative","Procedural","Subjective","Behavior")
stargazer(Indexes,summary=FALSE ,title="Questions included in each
Index for Analysis", colnames=FALSE, out =
"Rresults/indexquestions.html")

#Now we add a calculated column to each dataset with these composite
variables
#Calculated by the mean of the responses for the identified questions
#We omit any empty responses
PreIntervention <-
transform(PreIntervention,Declarative=round(rowMeans(PreIntervention[D
eclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(PreIntervention[ProceduralVec],na.rm=TRUE),1
),

Subjective=round(rowMeans(PreIntervention[SubjectiveVec],na.rm=TRUE),1
),

Behavior=round(rowMeans(PreIntervention[BehaviorVec],na.rm=TRUE),1))
PostIntervention <-
transform(PostIntervention,Declarative=round(rowMeans(PostIntervention
[DeclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(PostIntervention[ProceduralVec],na.rm=TRUE),
1),

Subjective=round(rowMeans(PostIntervention[SubjectiveVec],na.rm=TRUE),
1),

Behavior=round(rowMeans(PostIntervention[BehaviorVec],na.rm=TRUE),1))
FollowIntervention <-
transform(FollowIntervention,Declarative=round(rowMeans(FollowInterven
tion[DeclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(FollowIntervention[ProceduralVec],na.rm=TRUE
),1),
```

```
Subjective=round(rowMeans(FollowIntervention[SubjectiveVec],na.rm=TRUE
),1),

Behavior=round(rowMeans(FollowIntervention[BehaviorVec],na.rm=TRUE),1)
)
NonIntervention <-
transform(NonIntervention,Declarative=round(rowMeans(NonIntervention[D
eclarativeVec],na.rm=TRUE),1),

Procedural=round(rowMeans(NonIntervention[ProceduralVec],na.rm=TRUE),1
),

Subjective=round(rowMeans(NonIntervention[SubjectiveVec],na.rm=TRUE),1
),

Behavior=round(rowMeans(NonIntervention[BehaviorVec],na.rm=TRUE),1))

#Add variable for survey type
PreIntervention$Survey <-"Pre"
PostIntervention$Survey <- "Post"
FollowIntervention$Survey <- "Followup"
NonIntervention$Survey <- "Pre"

#We need to fix the factor levels so that all data frames have the
same set
Degree <- c("Administración Agropecuario","Desarrollo y Gestión
Interculturales","Economia Industrial","Fisioterapía","Odontología")
levels(PreIntervention$Degree) <-
c(levels(PreIntervention$Degree),"Economia Industrial")
levels(PostIntervention$Degree) <-
c(levels(PostIntervention$Degree),"Economia Industrial","Desarrollo y
Gestión Interculturales")
levels(FollowIntervention$Degree) <-
c(levels(FollowIntervention$Degree),"Economia Industrial","Desarrollo
y Gestión Interculturales")
levels(NonIntervention$Degree) <-
c(levels(NonIntervention$Degree),"Fisioterapía","Odontología")

PreIntervention$Degree <- factor(PreIntervention$Degree,levels=Degree)
PostIntervention$Degree <-
factor(PostIntervention$Degree,levels=Degree)
FollowIntervention$Degree <-
factor(FollowIntervention$Degree,levels=Degree)
FollowIntervention$Degree <-
factor(FollowIntervention$Degree,levels=Degree)

#Make Leon the base factor case
Origin <- c("Other","Leon","GTO","DF")
PreIntervention$Origin <- factor(PreIntervention$Origin,levels=Origin)
PostIntervention$Origin <-
factor(PostIntervention$Origin,levels=Origin)
```

```
FollowIntervention$Origin <-
factor(FollowIntervention$Origin,levels=Origin)
NonIntervention$Origin <- factor(NonIntervention$Origin,levels=Origin)

#Make Males the base factor
PreIntervention$Sex <- factor(PreIntervention$Sex,levels=c("m","f"))
PostIntervention$Sex <- factor(PostIntervention$Sex,levels=c("m","f"))
FollowIntervention$Sex <-
factor(FollowIntervention$Sex,levels=c("m","f"))
NonIntervention$Sex <- factor(NonIntervention$Sex,levels=c("m","f"))

#Create a new variable which groups the ages
PreIntervention$AgeCategory <- cut(PreIntervention$Age,c(0,23,30,100))
PostIntervention$AgeCategory <-
cut(PostIntervention$Age,c(0,23,30,100))
FollowIntervention$AgeCategory <-
cut(FollowIntervention$Age,c(0,23,30,100))
NonIntervention$AgeCategory <- cut(NonIntervention$Age,c(0,23,30,100))

#Instead We'll just change the failed grades from 0 to 5 so they don't
skew things quite so much
PreIntervention <- PreIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))
PostIntervention <- PostIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))
FollowIntervention <- FollowIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))
NonIntervention <- NonIntervention %>%
mutate(Final.Grade=replace(Final.Grade,Final.Grade==0,5))

#Removing outliers using Cook's Distance and boxplots
PreIntervention <- PreIntervention[-c(27,78),]
PostIntervention <- PostIntervention[-c(45),]
FollowIntervention <- FollowIntervention[-c(21),]
NonIntervention <- NonIntervention[-c(17,26),]


#Create dataframe with the pre- observations from all of the students
Allstudentsurvey <- rbind(PreIntervention,NonIntervention)

#Create matching pre/post/followup dataframes
#Pre/Post
Presurvey_pre_postmatch <-
semi_join(PreIntervention,PostIntervention,by="UniqueID")
Postsurvey_pre_postmatch <-
semi_join(PostIntervention,PreIntervention,by="UniqueID")
Presurvey_pre_postmatch <- arrange(Presurvey_pre_postmatch,UniqueID)
Postsurvey_pre_postmatch <- arrange(Postsurvey_pre_postmatch,UniqueID)

#Pre/Followup
Presurvey_pre_followupmatch <-
semi_join(PreIntervention,FollowIntervention,by="UniqueID")
```

```
Followupsurvey_pre_followupmatch <-
semi_join(FollowIntervention,PreIntervention,by="UniqueID")
Presurvey_pre_followupmatch <-
arrange(Presurvey_pre_followupmatch,UniqueID)
Followupsurvey_pre_followupmatch<-
arrange(Followupsurvey_pre_followupmatch,UniqueID)

#Post/Followup
Postsurvey_post_followupmatch <-
semi_join(PostIntervention,FollowIntervention,by="UniqueID")
Followupsurvey_post_followupmatch <-
semi_join(FollowIntervention,PostIntervention,by="UniqueID")
Postsurvey_post_followupmatch <-
arrange(Postsurvey_post_followupmatch,UniqueID)
Followupsurvey_post_followupmatch<-
arrange(Followupsurvey_post_followupmatch,UniqueID)

#For Hypothesis 3 and 4 we are going to create dataframes of the
differences of individuals
Differences_Pre_Post <- Postsurvey_pre_postmatch[,8:33]-
Presurvey_pre_postmatch[,8:33]
Differences_Pre_Post <-
cbind(Postsurvey_pre_postmatch[,c(1:7,35)],Differences_Pre_Post)
Differences_Pre_Followup <- Followupsurvey_pre_followupmatch[,8:33]-
Presurvey_pre_followupmatch[,8:33]
Differences_Pre_Followup <-
cbind(Followupsurvey_pre_followupmatch[,c(1:7,35)],Differences_Pre_Fol
lowup)
Differences_Post_Followup <- Postsurvey_post_followupmatch[,8:33]-
Followupsurvey_post_followupmatch[,8:33]
Differences_Post_Followup <-
cbind(Followupsurvey_post_followupmatch[,c(1:7,35)],Differences_Post_F
ollowup)
```

E3. Identify and remove outliers

```
#Doing Regressions on all the datasets to look at relationship of
Behavior to predictors
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
RegressPre <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PreIntervention)
RegressPost <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PostIntervention)
RegressFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = FollowIntervention)
```

```
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
RegressDiff_PreFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Followup)

#ID outliers
Allcooks <- data.frame(cooks.distance(RegressAll))
boxplot(Allcooks)
Precooks<- data.frame(cooks.distance(RegressPre))
boxplot(Precooks)
Postcooks<- data.frame(cooks.distance(RegressPost))
boxplot(Postcooks)
Followcooks<- data.frame(cooks.distance(RegressFollow))
boxplot(Followcooks)
Diffcooks <- data.frame(cooks.distance(RegressDiff))
boxplot(Diffcooks)

#Saving a JPEG of the boxplot included in the Appendix
jpeg("Rresults/AllStudentCooksBoxplot.jpg")
boxplot(Allcooks)
dev.off()

#The outliers we chose to remove
PreIntervention <- PreIntervention[-c(27,78),]
PostIntervention <- PostIntervention[-c(45),]
FollowIntervention <- FollowIntervention[-c(21),]
Allstudentsurvey <- Allstudentsurvey[-c(27,78,101,110),]
```

E4. Produce descriptive statistics of the samples

```
#We are going to create a couple of custom function to produce a
simple table of descriptive statistics
#Reports the precentages of variables in one column
Percentage <- function(x){
  tbl <- round(prop.table(table(x))*100,digits=1)
  res <- cbind(tbl)
  colnames(res) <- c("x")
  res
}

#Takes the mean of a variable
MyMean <- function(x){
  m <- round(mean(x,na.rm=TRUE),1)
  m
}

#Creates a single column of all the variables we want in our
descriptive table for a dataset
DescriptiveStats <- function(x){
```

```
  Prec <- do.call(rbind,lapply(x[c(2,4,5)],Percentage))
  full <- rbind(nrow(x),Prec,MyMean(x$Age),MyMean(x$Final.Grade))
  full
  }


#Creating the descriptive stats for each dataset
All <- DescriptiveStats(Allstudentsurvey)
Intervention <- DescriptiveStats(PreIntervention)
Followup <- DescriptiveStats(FollowIntervention)

#Creating one table which will be outputed as an html table
DescriptiveTable <- cbind(All,Intervention,Followup)
row.names(DescriptiveTable) <-
c("Number","Male","Female","Elsewhere","Leon","Guanajuato
State","Mexico City","Agricultural Administration","Intercultural
Development and Management","Industrial Economics","Physical
Therapy","Dentistry","Age(mean)","Grade(mean)")
colnames(DescriptiveTable) <- c("All","Intervention","Follow-up")
stargazer(DescriptiveTable,summary=FALSE,out="Rresults/DescriptiveStat
s.html",digits=1, title = "Descriptive Statistics of the Surveyed
Samples")
```

E5. Compare the samples statistically

```
#First we will compare the samples of students who participated in the
intervention and thos we didn't by using their scores on the pre-class
survey
#We are going to look at the index scores (Declarative, Procedural,
Subjective and Behavior)
#First we'll take the means of each sample
Intermeans <-
round(do.call(rbind,lapply(PreIntervention[,30:33],mean)),2)
Nonmeans <-
round(do.call(rbind,lapply(NonIntervention[,30:33],mean)),2)

#The Kolmogorov-Smirnov test can be used to test the null-hypothesis
that two samples come from the same population based on a variable
#We will reject the null-hypothesis if the p-value is <0.05 at which
point the statistic estimates the size of the difference
KSstat <-
data.frame(t(mapply(ks.test,PreIntervention[,30:33],NonIntervention[,3
0:33])))
KSstat <- select(KSstat,statistic,p.value)
KSstat <-
transform(KSstat,statistic=round(as.numeric(statistic),2),p.value=roun
d(as.numeric(p.value),2))

#We combine these two comparison approaches into one table and print
it in HTML
IntervsNon<- data.frame(Intermeans,Nonmeans,KSstat)
```

```
colnames(IntervsNon) <- c("Intervention (mean)","Nonintervention
(mean)","KS Statistic","KS P-Value")
stargazer(IntervsNon,summary=FALSE,out="Rresults/InterVsNonSamples.htm
l",digits=2,title = "KS Test comparing Intervention and
Nonintervention Samples")

#Second we are concerned that there might be a statistical difference
in the students who bothered to follow-up one year later from the
overall intervention group
#So we will compare the pre-class scores of students who followed up
with those that didnt't
#For that we need to create two new databases
NonFollowInter <-
anti_join(PreIntervention,FollowIntervention,by="UniqueID")
FollowInterPre <-
semi_join(PreIntervention,FollowIntervention,by="UniqueID")

#Then we can repeate what we did for our first comparison
KSstat2 <-
data.frame(t(mapply(ks.test,FollowInterPre[,30:33],NonFollowInter[,30:
33])))
KSstat2 <- select(KSstat2,statistic,p.value)
KSstat2 <-
transform(KSstat2,statistic=round(as.numeric(statistic),2),p.value=rou
nd(as.numeric(p.value),2))
NonFollowmeans <-
round(do.call(rbind,lapply(NonFollowInter[,30:33],mean)),2)
FollowPremeans <-
round(do.call(rbind,lapply(FollowInterPre[,30:33],mean)),2)
FollowvsNon<- data.frame(FollowPremeans,NonFollowmeans,KSstat2)
colnames(FollowvsNon) <- c("Followups (mean)","Non-Followups
(mean)","KS Statistic","KS P-Value")

KSstat3 <-
ks.test(FollowInterPre$Final.Grade,NonFollowInter$Final.Grade)
NonFollowmeans3 <- round(mean(NonFollowInter$Final.Grade),2)
Followmeans3 <- round(mean(FollowInterPre$Final.Grade),2)
GradeFollow <-
data.frame(Followmeans3,NonFollowmeans3,KSstat3$statistic,KSstat3$p.va
lue)
colnames(GradeFollow) <- c("Followups (mean)","Non-Followups
(mean)","KS Statistic","KS P-Value")
row.names(GradeFollow) <- "Final Grade"

FollowvsNon <- rbind(FollowvsNon,GradeFollow)

stargazer(FollowvsNon,summary=FALSE,out="Rresults/FollowVsNonSamples.h
tml",digits=2,title = "KS Test comparing Preintervention sample with
Followup")
```

E6. Assessing the indices with Cronbach's Alpha

```
#Going to take a look at the indexes of the knowledge domains and
behavior for the all student survey
#First we'll calculate the mean and standard deviation for each
Allmeans <-
round(do.call(rbind,lapply(Allstudentsurvey[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(Allstudentsurvey[,30:33],sd)),2)

#To assess the index's reliability of measure we calculate Cronbach's
Alpha and report the standardized number (1 being all items are 100%
correlated)
DecA <- alpha(Allstudentsurvey[,DeclarativeVec])
DecP <- alpha(Allstudentsurvey[,ProceduralVec])
DecS <- alpha(Allstudentsurvey[,SubjectiveVec])
DecB <- alpha(Allstudentsurvey[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)

#Create the html table
DomainsAll <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsAll) <- c("Mean","Standard Deviation", "Standardized
Alpha")
stargazer(DomainsAll,summary=FALSE,out="Rresults/DomainsAllAlpha.html"
,digits=2)

#Pre
Allmeans <-
round(do.call(rbind,lapply(PreIntervention[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(PreIntervention[,30:33],sd)),2)
DecA <- alpha(PreIntervention[,DeclarativeVec])
DecP <- alpha(PreIntervention[,ProceduralVec])
DecS <- alpha(PreIntervention[,SubjectiveVec])
DecB <- alpha(PreIntervention[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)
DomainsPre <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsPre) <- c("Mean","Standard Deviation", "Standardized
Alpha")
stargazer(DomainsPre,summary=FALSE,out="Rresults/DomainsPreAlpha.html"
,digits=2)

#Post
Allmeans <-
round(do.call(rbind,lapply(PostIntervention[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(PostIntervention[,30:33],sd)),2)
DecA <- alpha(PostIntervention[,DeclarativeVec])
```

```
DecP <- alpha(PostIntervention[,ProceduralVec])
DecS <- alpha(PostIntervention[,SubjectiveVec])
DecB <- alpha(PostIntervention[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)
DomainsPost <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsPost) <- c("Mean","Standard Deviation", "Standardized
Alpha")
stargazer(DomainsPost,summary=FALSE,out="Rresults/DomainsPostAlpha.htm
l",digits=2)

#Follow
Allmeans <-
round(do.call(rbind,lapply(FollowIntervention[,30:33],mean)),2)
Allsd <- round(do.call(rbind,lapply(FollowIntervention[,30:33],sd)),2)
DecA <- alpha(FollowIntervention[,DeclarativeVec])
DecP <- alpha(FollowIntervention[,ProceduralVec])
DecS <- alpha(FollowIntervention[,SubjectiveVec])
DecB <- alpha(FollowIntervention[,BehaviorVec])
Allalpha <-
t(data.frame(DecA$total[2],DecP$total[2],DecS$total[2],DecB$total[2]))
Allalpha <- transform(Allalpha,Alpha=round(as.numeric(X_data),2))
Allalpha <- select(Allalpha,Alpha)
DomainsFollow <- data.frame(Allmeans,Allsd,Allalpha)
colnames(DomainsFollow) <- c("Mean","Standard Deviation",
"Standardized Alpha")
stargazer(DomainsFollow,summary=FALSE,out="Rresults/DomainsFollowAlpha
.html",digits=2)
```

E7. Multicollinearity among independent variables in the regressions

```
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
RegressPre <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PreIntervention)
RegressPost <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PostIntervention)
RegressFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = FollowIntervention)
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
RegressDiff_PreFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Followup)
```

```
#Check for colinearity among regressors by calculating Variance
Influence Factors
VarianceInfluenceFactors <-
data.frame(vif(RegressAll),vif(RegressPre),vif(RegressPost),vif(Regres
sFollow),vif(RegressDiff),vif(RegressDiff_PreFollow))
VarianceInfluenceFactors <-
dplyr::select(VarianceInfluenceFactors,GVIF,GVIF.1,GVIF.2,GVIF.3,GVIF.
4,GVIF.5)
colnames(VarianceInfluenceFactors) <- c("All Students","Pre
Intervention","Post Intervention","Followup","Differences
Pre/Post","Differences Pre/Followup")
stargazer(VarianceInfluenceFactors,summary=FALSE,
out="Rresults/VIF.html",title = "Variance Influence Factors for all
Regressions")

#The VIFs are all low except for the Follow-up sample so we need to be
very cautious about interpreting those results
#For the follow-up sample the three knowledge indices appear to be
tightly correlated with each other
```

E8. Table of All Regression Results

```
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
StdRegressAll <- standardize(RegressAll)
RegressPre <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PreIntervention)
StdRegressPre <- standardize(RegressPre)
RegressPost <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = PostIntervention)
StdRegressPost <- standardize(RegressPost)
RegressFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = FollowIntervention)
StdRegressFollow <- standardize(RegressFollow)
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
StdRegressDiff <- standardize(RegressDiff)
RegressDiff_PreFollow <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Followup)
StdRegressDiff_PreFollow <- standardize(RegressDiff_PreFollow)

stargazer(RegressAll,RegressPre,RegressPost,RegressFollow,RegressDiff,
RegressDiff_PreFollow, out = "Rresults/AllRegressions.html",
column.labels = c("All Students","Pre Intervention","Post
```

```
Intervention","Followup","Differences Pre/Post","Differences
Pre/Followup"),title = "Summary of all regression results in study")
stargazer(StdRegressAll,StdRegressPre,StdRegressPost,StdRegressFollow,
StdRegressDiff,StdRegressDiff_PreFollow, out =
"Rresults/AllRegressionsStandardized.html", column.labels = c("All
Students","Pre Intervention","Post
Intervention","Followup","Differences Pre/Post","Differences
Pre/Followup"),title = "Summary of standardized regression results in
study")
```

E9. Evaluating Hypothesis 1

```
#Do regression on dataset of all-students to look at relationship of
Behavior and Indices
RegressAll <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Allstudentsurvey)
StdRegressAll <- (RegressAll)
stargazer(StdRegressAll,out =
"Rresults/H1_AllStudentSurveyRegression.html", column.labels =
"Standardized All Students",title = "Hypothesis 1: Regression
Examining Relationship between Knowledge and Behavior")
```

E10. Evaluating Hypothesis 2

```
#We conduct a paired T-Test comparing each students pre and post
intervention survey responses
Ttest_PrePostQs <-
mapply(t.test,Postsurvey_pre_postmatch[,8:29],Presurvey_pre_postmatch[
,8:29],paired=TRUE, conf.level=0.95)
Ttest_PrePostQs <- data.frame(Ttest_PrePostQs)
Ttest_PrePostQs <- data.frame(t(Ttest_PrePostQs))
Ttest_PrePostQs[,c(6,7,8,9)] <- list(NULL)
Ttest_PrePostQs <- cSplit(Ttest_PrePostQs,"conf.int")
Ttest_PrePostQs <-
transform(Ttest_PrePostQs,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PrePostQs[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
Ttest_PrePostQs <-
mutate(Ttest_PrePostQs,Adjusted.P.Values=p.adjust(p.value,method="BY")
)
Ttest_PrePostQs$Names <- xaxis_names[1:22]
Ttest_PrePostQs$Names <-
factor(Ttest_PrePostQs$Names,levels=Ttest_PrePostQs$Names)
#Effect Size with Cohen's D
```

```
CohensDEffect_PrePostQs <-
mapply(cohen.d,Postsurvey_pre_postmatch[,8:29],Presurvey_pre_postmatch
[,8:29],paired=TRUE)
CohensDEffect_PrePostQs <- data.frame(CohensDEffect_PrePostQs)
CohensDEffect_PrePostQs <- data.frame(t(CohensDEffect_PrePostQs))
CohensDEffect_PrePostQs[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PrePostQs <- cSplit(CohensDEffect_PrePostQs,"conf.int")
CohensDEffect_PrePostQs <-
transform(CohensDEffect_PrePostQs,CohensD=as.numeric(estimate),CD.conf
.int_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric(c
onf.int_2))
CohensDEffect_PrePostQs[,c(1,2,3)] <- list(NULL)
TtestCohensD_PrePostQs <-
data.frame(Ttest_PrePostQs,CohensDEffect_PrePostQs)
TtestCohensD_PrePostQs <-
TtestCohensD_PrePostQs[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PrePostQs$Names <-
factor(TtestCohensD_PrePostQs$Names,levels=TtestCohensD_PrePostQs$Name
s)
#Printing the Results to HTML
stargazer(TtestCohensD_PrePostQs,title="T-test Results and Cohen's D
for Pre vs. Post Change in Survey",
summary=FALSE,out="Rresults/H2_PrePostTtestAllQs.html")
#Graphing the results and saving
ggplot(TtestCohensD_PrePostQs, aes(x=Names, y=estimate,
ymin=conf.int_min,ymax=conf.int_max),ordered=FALSE) +
geom_pointrange() + theme_bw() + coord_flip() + geom_hline(yintercept
= 0,linetype = "dotted") + ylab("T test Estimate of Mean Response
Change")+xlab("")
ggsave("Rresults/H2_GraphPrePostAllQs_Vertical.jpg")


#Ttest for just the indices which is another way of dealing with
multiple comparisons
Ttest_PrePost <-
mapply(t.test,Postsurvey_pre_postmatch[,30:33],Presurvey_pre_postmatch
[,30:33],paired=TRUE, conf.level=0.95)
Ttest_PrePost <- data.frame(Ttest_PrePost)
Ttest_PrePost <- data.frame(t(Ttest_PrePost))
Ttest_PrePost[,c(6,7,8,9)] <- list(NULL)
Ttest_PrePost <- cSplit(Ttest_PrePost,"conf.int")
Ttest_PrePost <-
transform(Ttest_PrePost,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PrePost[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
```

```
Ttest_PrePost <-
mutate(Ttest_PrePost,Adjusted.P.Values=p.adjust(p.value,method="BY"))
Ttest_PrePost$Names <-
c("Declarative","Procedural","Subjective","Behavior")
Ttest_PrePost$Names <-
factor(Ttest_PrePost$Names,levels=Ttest_PrePost$Names)
#Effect Size with Cohen's D
CohensDEffect_PrePost <-
mapply(cohen.d,Postsurvey_pre_postmatch[,30:33],Presurvey_pre_postmatc
h[,30:33],paired=TRUE)
CohensDEffect_PrePost <- data.frame(CohensDEffect_PrePost)
CohensDEffect_PrePost <- data.frame(t(CohensDEffect_PrePost))
CohensDEffect_PrePost[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PrePost <- cSplit(CohensDEffect_PrePost,"conf.int")
CohensDEffect_PrePost <-
transform(CohensDEffect_PrePost,CohensD=as.numeric(estimate),CD.conf.i
nt_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric(con
f.int_2))
CohensDEffect_PrePost[,c(1,2,3)] <- list(NULL)
TtestCohensD_PrePostQs <-
data.frame(Ttest_PrePost,CohensDEffect_PrePost)
TtestCohensD_PrePostQs <-
TtestCohensD_PrePostQs[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PrePostQs$Names <-
factor(TtestCohensD_PrePostQs$Names,levels=TtestCohensD_PrePostQs$Name
s)
#Printing the Results to HTML
stargazer(TtestCohensD_PrePostQs,title="T-test Results and Cohen's D
for Pre vs. Post Change in Indices",
summary=FALSE,out="Rresults/H2_PrePostTtestIndices.html")
#Graphing the results and saving
ggplot(TtestCohensD_PrePostQs, aes(x=Names, y=CohensD,
ymin=CD.conf.int_min,ymax=CD.conf.int_max),ordered=FALSE) +
geom_pointrange(size=2,shape=15) +
theme_bw()+theme(text=element_text(size=20)) + coord_flip() +
geom_hline(yintercept = 0,linetype = "longdash",size=1) + ylab(""Cohen
D Effect Size Estimate"")+xlab("")
ggsave("Rresults/H2_GraphPrePostIndices_Vertical.jpg")
```

E11. Evaluating Hypothesis 3

```
#Data frame created with the differences between pre and post for
individuals
#Run the same regression as for H1
RegressDiff <-
lm(Behavior~Sex+AgeCategory+Origin+Degree+Final.Grade+Declarative+Proc
edural+Subjective,data = Differences_Pre_Post)
STDRegressDiff <- standardize(RegressDiff)
stargazer(STDRegressDiff,out="Rresults/H3_RegressionOfDifferences.html
",column.labels = "Difference Between Pre/Post",title = "Hypothesis 3:
Relationship between the change in knowledge and change in behavior")
```

E12. Evaluating Hypothesis 4

```
#Main test of hypothesis is to see if t-test shows difference between
pre and followup
Ttest_PreFollowQs <-
mapply(t.test,Followupsurvey_pre_followupmatch[,8:29],Presurvey_pre_fo
llowupmatch[,8:29],paired=TRUE, conf.level=0.95)
Ttest_PreFollowQs <- data.frame(Ttest_PreFollowQs)
Ttest_PreFollowQs <- data.frame(t(Ttest_PreFollowQs))
Ttest_PreFollowQs[,c(6,7,8,9)] <- list(NULL)
Ttest_PreFollowQs <- cSplit(Ttest_PreFollowQs,"conf.int")
Ttest_PreFollowQs <-
transform(Ttest_PreFollowQs,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PreFollowQs[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
Ttest_PreFollowQs <-
mutate(Ttest_PreFollowQs,Adjusted.P.Values=p.adjust(p.value,method="BY
"))
Ttest_PreFollowQs$Names <- xaxis_names[1:22]
Ttest_PreFollowQs$Names <-
factor(Ttest_PreFollowQs$Names,levels=Ttest_PreFollowQs$Names)
#Effect Size with Cohen's D
CohensDEffect_PreFollowQs <-
mapply(cohen.d,Followupsurvey_pre_followupmatch[,8:29],Presurvey_pre_f
ollowupmatch[,8:29],paired=TRUE)
CohensDEffect_PreFollowQs <- data.frame(CohensDEffect_PreFollowQs)
CohensDEffect_PreFollowQs <- data.frame(t(CohensDEffect_PreFollowQs))
CohensDEffect_PreFollowQs[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PreFollowQs <-
cSplit(CohensDEffect_PreFollowQs,"conf.int")
CohensDEffect_PreFollowQs <-
transform(CohensDEffect_PreFollowQs,CohensD=as.numeric(estimate),CD.co
nf.int_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric
(conf.int_2))
CohensDEffect_PreFollowQs[,c(1,2,3)] <- list(NULL)
TtestCohensD_PreFollowQs <-
data.frame(Ttest_PreFollowQs,CohensDEffect_PreFollowQs)
TtestCohensD_PreFollowQs <-
TtestCohensD_PreFollowQs[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PreFollowQs$Names <-
factor(TtestCohensD_PreFollowQs$Names,levels=TtestCohensD_PreFollowQs$
Names)
#Printing the Results to HTML
```

```
stargazer(TtestCohensD_PreFollowQs,title="T-test Results and Cohen's D
for Pre vs. Follow Change in Survey",
summary=FALSE,out="Rresults/H4_PreFollowTtestAllQs.html")
#Graphing the results and saving
ggplot(TtestCohensD_PreFollowQs, aes(x=Names, y=estimate,
ymin=conf.int_min,ymax=conf.int_max),ordered=FALSE) +
geom_pointrange() + theme_bw() + coord_flip() + geom_hline(yintercept
= 0,linetype = "dotted") + ylab("T test Estimate of Mean Response
Change")+xlab("")
ggsave("Rresults/H4_GraphPreFollowAllQs_Vertical.jpg")


#Ttest for just the indices which is another way of dealing with
multiple comparisons
Ttest_PreFollow <-
mapply(t.test,Followupsurvey_pre_followupmatch[,30:33],Presurvey_pre_f
ollowupmatch[,30:33],paired=TRUE, conf.level=0.95)
Ttest_PreFollow <- data.frame(Ttest_PreFollow)
Ttest_PreFollow <- data.frame(t(Ttest_PreFollow))
Ttest_PreFollow[,c(6,7,8,9)] <- list(NULL)
Ttest_PreFollow <- cSplit(Ttest_PreFollow,"conf.int")
Ttest_PreFollow <-
transform(Ttest_PreFollow,Tstatistic=as.numeric(statistic),
DegreesFreedom=as.numeric(parameter),
p.value=as.numeric(p.value),estimate=as.numeric(estimate),

conf.int_min=extract_numeric(conf.int_1),conf.int_max=extract_numeric(
conf.int_2))
Ttest_PreFollow[,c(1,2,5,6)] <- list(NULL)
#Adjust p-values for multiple comparisons using Benjamini & Yekkutieli
(2001)
Ttest_PreFollow <-
mutate(Ttest_PreFollow,Adjusted.P.Values=p.adjust(p.value,method="BY")
)
Ttest_PreFollow$Names <-
c("Declarative","Procedural","Subjective","Behavior")
Ttest_PreFollow$Names <-
factor(Ttest_PreFollow$Names,levels=Ttest_PreFollow$Names)
#Effect Size with Cohen's D
CohensDEffect_PreFollow <-
mapply(cohen.d,Followupsurvey_pre_followupmatch[,30:33],Presurvey_pre_
followupmatch[,30:33],paired=TRUE)
CohensDEffect_PreFollow <- data.frame(CohensDEffect_PreFollow)
CohensDEffect_PreFollow <- data.frame(t(CohensDEffect_PreFollow))
CohensDEffect_PreFollow[,c(1,2,5,6)] <- list(NULL)
CohensDEffect_PreFollow <- cSplit(CohensDEffect_PreFollow,"conf.int")
CohensDEffect_PreFollow <-
transform(CohensDEffect_PreFollow,CohensD=as.numeric(estimate),CD.conf
.int_min=extract_numeric(conf.int_1),CD.conf.int_max=extract_numeric(c
onf.int_2))
CohensDEffect_PreFollow[,c(1,2,3)] <- list(NULL)
```

```
TtestCohensD_PreFollow <-
data.frame(Ttest_PreFollow,CohensDEffect_PreFollow)
TtestCohensD_PreFollow <-
TtestCohensD_PreFollow[,c(8,2,1,7,5,6,3,4,9,10,11)]
TtestCohensD_PreFollow$Names <-
factor(TtestCohensD_PreFollow$Names,levels=TtestCohensD_PreFollow$Name
s)
#Printing the Results to HTML
stargazer(TtestCohensD_PreFollow,title="T-test Results and Cohen's D
for Pre vs. Followup Change in Indices",
summary=FALSE,out="Rresults/H4_PreFollowupTtestIndices.html")
#Graphing the results and saving
ggplot(TtestCohensD_PreFollow, aes(x=Names, y=CohensD,
ymin=CD.conf.int_min,ymax=CD.conf.int_max),ordered=FALSE) +
geom_pointrange(size=2,shape=15) +
theme_bw()+theme(text=element_text(size=20)) + coord_flip() +
geom_hline(yintercept = 0,linetype = "longdash",size=1) + ylab("Cohen
D Effect Size Estimate")+xlab("")
ggsave("Rresults/H4_GraphPreFollowIndices_Vertical.jpg")
```

E13. Density Plot Histograms of the Indices

```
#Denisty Plot Histograms of indexes
Alldata <- rbind(Allstudentsurvey,PostIntervention,FollowIntervention)
Alldata$Survey <-
factor(Alldata$Survey,levels=c("Pre","Post","Followup"))

ggplot(Alldata,aes(x=Behavior,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotBehavior.jpeg")
ggplot(Alldata,aes(x=Declarative,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotDeclarative.jpeg")
ggplot(Alldata,aes(x=Procedural,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotProcedural.jpeg")
ggplot(Alldata,aes(x=Subjective,
fill=Survey))+geom_density(alpha=.3)+xlim(1,4)+ylim(0,1.3)+ylab("Frequ
ency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotSubjective.jpeg")
```

E14. Investigation into the Sex Differences

```
#First we are create new databases of just the female and just the
male students
All_Female <- Allstudentsurvey %>% filter(Sex=="f")
All_Male <- Allstudentsurvey %>% filter(Sex=="m")
Post_Female <- PostIntervention %>% filter(Sex=="f")
Post_Male <- PostIntervention %>% filter(Sex=="m")
Follow_Female <- FollowIntervention %>% filter(Sex=="f")
Follow_Male <- FollowIntervention %>% filter(Sex=="m")

#Ran OLs regressions on each sex
RegressAll_Female <-
lm(Behavior~AgeCategory+Origin+Degree+Final.Grade+Declarative+Procedur
al+Subjective,data = All_Female)
RegressAll_Male <-
lm(Behavior~AgeCategory+Origin+Degree+Final.Grade+Declarative+Procedur
al+Subjective,data = All_Male)
stargazer(RegressAll_Female,RegressAll_Male,
out="Rresults/SexRegressions.html", column.labels =
c("Female","Male"), title = "Regressions of All Student Survey
Seperated by Sex")

STDRegressAll_Female <- standardize(RegressAll_Female)
STDRegressAll_Male <- standardize(RegressAll_Male)
stargazer(STDRegressAll_Female,STDRegressAll_Male,
out="Rresults/SexRegressionsSTD.html", column.labels =
c("Female","Male"), title = "Standardized Regressions of All Student
Survey Seperated by Sex")

#Ttests comparing male and female sustainable behaviors for each of
the three samples
#Then lots of work to get it in a usable format
Ttest_Sex_All <- matrix(t.test(All_Female$Behavior,All_Male$Behavior))
Ttest_Sex_Post <-
matrix(t.test(Post_Female$Behavior,Post_Male$Behavior))
Ttest_Sex_Follow <-
matrix(t.test(Follow_Female$Behavior,Follow_Male$Behavior))
Ttest_Sex <- data.frame(Ttest_Sex_All,Ttest_Sex_Post,Ttest_Sex_Follow)
Ttest_Sex <- data.frame(t(Ttest_Sex))
Ttest_Sex[,c(6,7,8,9)] <- list(NULL)
Ttest_Sex <- cSplit(Ttest_Sex,c("X4","X5"))
Ttest_Sex <-
transmute(Ttest_Sex,t.statistic=as.numeric(X1),parameter=as.numeric(X2
),p.value=as.numeric(X3),FemaleMean=extract_numeric(X5_1),MaleMean=ext
ract_numeric(X5_2),DiffofMeans=FemaleMean-MaleMean,
confint_min=extract_numeric(X4_1),confint_max=extract_numeric(X4_2))
Ttest_Sex$Surveys <- c("All","Post","Followup")
Ttest_Sex$Surveys <-
factor(Ttest_Sex$Surveys,levels=Ttest_Sex$Surveys)
```

```
ggplot(Ttest_Sex, aes(x=Surveys, y=DiffofMeans,
ymin=confint_min,ymax=confint_max),ordered=FALSE) +
geom_pointrange(size=2,shape=15) +
theme_bw()+theme(text=element_text(size=20)) + coord_flip() +
geom_hline(yintercept = 0,linetype = "longdash",size=1) +
ylab("Difference Between the Sexes")+xlab("")
ggsave("Rresults/GraphofSexDiff.jpeg")
stargazer(Ttest_Sex,out="Rresults/Sexttest.html",summary=FALSE)

#Another way to visualize the difference:
ggplot(Allstudentsurvey,aes(x=Behavior, group=Sex,
fill=Sex))+geom_density(alpha=.3)+xlim(1,4)+ylab("Frequency of Index
Scores")+theme(text=element_text(size=20),legend.position="bottom",leg
end.title=element_blank())
ggsave("Rresults/DensityPlotSexAll.jpeg")

#KS Test:
KSstatSex <-
data.frame(t(mapply(ks.test,All_Female[,30:33],All_Male[,30:33])))
KSstatSex <- dplyr::select(KSstatSex,statistic,p.value)
KSstatSex <-
transform(KSstatSex,statistic=round(as.numeric(statistic),2),p.value=r
ound(as.numeric(p.value),2))
Femalemeans <- round(do.call(rbind,lapply(All_Female[,30:33],mean)),2)
Malemeans <- round(do.call(rbind,lapply(All_Male[,30:33],mean)),2)
FemaleVsMale<- data.frame(Femalemeans,Malemeans,KSstatSex)
colnames(FemaleVsMale) <- c("Female (mean)","Male (mean)","KS
Statistic","KS P-Value")
stargazer(FemaleVsMale,summary=FALSE,out="Rresults/KSSexComparison.htm
l",digits=2,title = "KS Test comparing Females vs Males for All
student Sample")
```

Appendix F: References in Appendices

Benjamini, Y., & Yekutieli, D. (2016). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, *29*(4), 1165–1188. http://doi.org/10.1214/aos/1013699998

Bland, J. M., & Altman, D. G. (1997). Statistics Notes: Cronbach's Alpha. *British Medical Journal*, *314*(7080), 22–23.

Brundiers, K., & Wiek, A. (2013). Do We Teach What We Preach? An International Comparison of Problem- and Project-Based Learning Courses in Sustainability. *Sustainability*, *5*(4), 1725–1746. http://doi.org/10.3390/su5041725

Chao, Y.-L., & Lam, S.-P. (2011). Measuring Responsible Environmental Behavior: Self-Reported and Other-Reported Measures and Their Differences in Testing a Behavioral Model. *Environment and Behavior*, *43*(1), 53–71. http://doi.org/10.1177/0013916509350849

Christensen, R., & Knezek, G. (2015). The climate change attitude survey: Measuring middle school student beliefs and intentions to enact positive environmental change. *International Journal of Environmental and Science Education*, *10*(5), 773–788. http://doi.org/10.12973/ijese.2015.276a

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, New Jersey: L. Erlbaum.

De Young, R. (2000). Expanding and Evaluating Motives for Environmentally Responsible Behavior. *Journal of Social Issues*, *56*(3), 509–526. http://doi.org/10.1111/0022-4537.00181

Frisk, E., & Larson, K. L. (2011). Educating for Sustainability: Competencies & Practices for Transformative Action. *Journal of Sustainability Education*, *2*(March). Retrieved from http://www.jsedimensions.org/wordpress/wp-content/uploads/2011/03/FriskLarson2011.pdf

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*, 2865–2873. http://doi.org/10.1002/sim.3605

Gossard, M. M. H., & York, R. (2003). Social structural influences on meat consumption. *Human Ecology Review*, *10*(1), 1–9. Retrieved from http://www.humanecologyreview.org/pastissues/her101/101gossardyork.pdf

Kim, C., & Storer, B. E. (1996). Reference values for cook's distance. *Communications in Statistics Part B: Simulation and Computation*, *25*(3), 691–708. http://doi.org/10.1080/03610919608813337

McKenzie-Mohr, D. (2011). *Fostering Sustainable Behavior: An Introduction to Community-Based Social Marketing*. New Society Publishers.

Milfont, T. L. (2009). The effects of social desirability on self-reported environmental attitudes and ecological behaviour. *Environmentalist*, *29*, 263–269. http://doi.org/10.1007/s10669-008-9192-2

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690. http://doi.org/10.1007/s11135-006-9018-6

Redman, E., & Redman, A. (2014). Transforming sustainable food and waste behaviors by realigning domains of knowledge in our education system. *Journal of Cleaner Production*, *64*, 147–157. http://doi.org/10.1016/j.jclepro.2013.09.016

Rogers, E. (2003). *Diffusion of innovations*. Simon and Schuster. Retrieved from

http://books.google.com/books?hl=en&lr=&id=v1ii4QsB7jIC&oi=fnd&pg=PR15&dq=Diffusion+of+innovations&ots=DJZvsRTl9O&sig=8OO6yzNGa9ElKzF1KKbmRs2D_9Y

Sahin, E., Ertepinar, H., & Teksoz, G. (2012). University students' behaviors pertaining to sustainability : A structural equation model with sustainability-related attributes. *International Journal of Environmental & Science Education*, *7*(3), 459–478.

Schochet, P. Z. (2008). *Guidelines for Multiple Testing in Impact Evaluations of Educational Interventions. Mathematica Policy Research, Inc.* Retrieved from http://eric.ed.gov/?id=ED502199