Comparing Retrospective and Pre-Post Survey Measurement of First-Year African Engineering Students' Skill and Perception Changes through a PBL Course

Yohance Lewis
Massachusetts Institute of Technology, USA, ylfl@mit.edu
Sampson Kofi Nani
Ashesi University, Ghana, sampson.nani@alumni.ashesi.edu.gh
Charity Ampomah
Ashesi University, Ghana, charity.ampomah@alumni.ashesi.edu.gh
Dr. Heather Beem
Ashesi University, Ghana, hbeem@ashesi.edu.gh

Abstract

For many first-year engineering students at Ashesi University in Ghana, the project-based learning (PBL) approach they experience in Introduction to Engineering differs from the rote-based pedagogies they have previously experienced. The authors' previous efforts to understand this pedagogy's impact on the students have relied on pre-post survey tools. For some responses, higher-than-expected scores on the pre-survey have been noticed. This study aimed to understand whether a different methodology, a retrospective pre-post survey, would yield similar results and what reasons may underlie any differences.

A mixed-method approach was deployed with students in the 2024 cohort of the course. They completed two surveys – one at the beginning (pre-survey) and one at the end of the course (retrospective pre-post survey). Five skill/perception categories were deployed with Likert-scale questions. T. tests revealed no statistically significant difference in the traditional pre-post analysis, while in the retrospective analysis, all five categories had statistically significant increases with small to medium effect sizes. Interviews with select respondents revealed self-assessment bias, as they tended to overestimate their skills at the beginning of the course. These results suggest that the retrospective approach may be better suited for studies on this target group.

Keywords: Retrospective survey, first-year, self-assessment bias, Dunning-Kruger effect, memory/recall bias

1 Introduction

1.1 Previous work with this population

Over the last few years, the authors have investigated the impact of project-based learning (PBL) on first-year students in an Introduction to Engineering class at their university - a small private institution in Ghana. These studies have sought to understand the students' experience, especially as many of them

encounter a PBL approach for the first time in this course. The studies have looked at how constructs such as self-efficacy are impacted (Beem, 2021a; Beem et al., 2023a; Ampomah & Beem, 2024), whether the physical environment plays a role (Beem, 2022), and whether this impact was experienced in an online format (Beem et al., 2023b). The methodological approach for conducting these investigations has been the use of a traditional pre-post survey.

Past work has focused on self-efficacy as the dominant construct in view, but it is worth expanding the scope to identify other constructs that may capture any impact the course has on these students. It is worthwhile to investigate a broader set of skills and perceptions that the students may increase or change through this course.

1.2 Survey types

The pre-post survey methodology is the traditional way of measuring the impact of interventions through self-assessments and has therefore been utilized in many studies (Davis et al., 2018; Ferguson et al., 2012; Hufford, 2010). This approach compares self-reports of perceptions at the beginning of an intervention with self-reports soon after the intervention (Kowalski, 2023). This is a highly popular method due to the straightforward nature of its deployment and analysis. This method could, however, be considered somewhat time-consuming and may result in respondent fatigue since the survey is conducted multiple times (Porter et al., 2004). When the author administered pre-surveys to this student population in the past, higher-than-expected scores were observed in some instances (Beem 2021a; Beem, 2021b). This was surprising since most students stated they had limited experience with the topics to be covered in class. This observation has motivated the exploration of other measurement approaches. An alternative means of measuring the impact of interventions through self-assessment is the retrospective pre-post survey. In this model, the post-test and pre-test are administered simultaneously at the end of the intervention through a single survey. The respondent is asked to rate themselves at both timepoints by reflecting on where they started as well as where they are now. This reduces the number of surveys that the respondent is asked to complete and can offer them an opportunity to more accurately reflect on the learnings gained. Both the traditional pre-post and the retrospective pre-post surveys have their vulnerabilities (Lamb, 2005; Nimon, 2014). Kowalski (2023), therefore, advises evaluators to determine which design is most likely to provide the most accurate results. This study incorporates a qualitative phase to further explore the causes of any inconsistencies between the two methods.

1.3 Participant biases

A commonly cited limitation of self-assessment surveys is their vulnerability to response-shift bias (Drennan & Hyde, 2008; Hill & Betz, 2005). The response-shift bias refers to the changes in internal standards of evaluation (Howard et al., 1979). The response-shift bias may arise from changes in participants' levels of self-awareness before or after an intervention. Response-shift is more likely if the intervention being administered increases a participant's awareness of the changes being measured (Moore & Tananis, 2009). In addition to the response-shift bias, several other biases may also be at play.

Another commonly cited bias is the Dunning-Kruger effect. This applies when individuals who are either underperformers or overperformers tend to underestimate or overestimate their abilities, respectively (Dunning, 2011). Hence, this bias causes participants to inaccurately assess their own capabilities. Similar biases include impact bias, self-serving bias, and overconfidence bias. Impact bias refers to the tendency to overestimate emotional responses to events and experiences (Grimes et al., 2015). This bias leads individuals to assume that specific tasks will be easier to perform without considering the task's long-term

emotional impact. The overconfidence bias also refers to the tendency to believe one is better than one actually is. While this bias can boost morale and ambition, it may also result in faulty assessments, unrealistic expectations, and poor decision-making (Fowler & Johnson, 2011). The self-serving bias, as defined by Forsyth (2008), is the inclination to attribute positive outcomes to internal factors and negative outcomes to external circumstances. This bias makes individuals absolve themselves from the negative effects of their actions and credit themselves for positive outcomes.

Another type of bias is the hindsight bias. Also known as the *"knew-it-all-along"* effect, this phenomenon leads individuals to believe that they should have anticipated the outcomes of an intervention after the results are known. Hindsight bias can be influenced by fleeting emotions (Zurbriggen et al., 2021). This positive anticipation can create a false sense of awareness in individuals, causing them to make inaccurate self-evaluations.

A final type of bias is the influence/response bias. This captures the tendency of survey respondents to alter their answers in order to align them with socially acceptable norms (Wetzel et al., 2016). Paulhaus (1991) also defines influence/response bias as the "systematic tendency to respond to a range of questionnaire items based on factors other than the specific content of the items" (p. 17). There are several types of influence/response biases, which are detailed in Table A1 of the Appendix, as posited by Wetzel et al. (2016). Within these types, Meade and Craig (2012) identified respondents' interest and the length of the survey as major factors. For longer surveys, it is reasonable to expect that respondents' attention may wane over time, leading to an increased likelihood of random responses, especially towards the middle or end of the survey (Baer et al., 1997; Berry et al., 1992).

There is a need to examine the influence of various biases on the survey responses provided by students in the authors' context. Hence, this paper seeks to answer the questions;

1) What transformational changes beyond self-efficacy are seen in first-year engineering students in a Project-based Learning (PBL) course?

2) How do pretests and retrospective tests compare in the design of surveys for a PBL course?

2 Methodology

This mixed-method study leveraged tools to conduct both quantitative and qualitative analyses. The two types of surveys were first administered to ascertain any differences in what the participants self-reported. Then, interviews were conducted with select participants to elucidate the reasons for any differences. This research was conducted with the student cohort that participated in Ashesi's Introduction to Engineering course from January to April 2024. Ethical clearance was received from Ashesi University's Institutional Review Board (IRB), with application number 1282023**,** which was valid until May 2025. To ensure informed consent, participants received detailed information about the study and were informed that they could opt out anytime. The invitation email for scheduling interviews emphasized that participation was voluntary and that they could withdraw without consequences. During the interview meeting, the researcher once again reminded participants of the voluntary nature of the research and sought their consent before proceeding with the questions. Labels were used to represent the participants instead of their names to ensure their anonymity. No other identifying information was used. The interview data is stored securely online, with access shared only with the author team.

2.1 Surveys

Two surveys were designed and deployed for the quantitative analysis: a pre-survey and a retrospective pre-post study. Both surveys asked participants to assess themselves along five categories, using the same Likert-scale questions (scale of 1 to 10). The extreme scores of 1 and 10 indicated "strongly disagree" and "strongly agree", respectively. These categories included students' perceptions of their Group Work Skills, Interpersonal Skills, Problem-Solving Skills, and Potential Within Africa and Self, and other questions. The full list of questions is provided in Table A2.

These questions had been formed as a result of interviews with this student population, which sought to elucidate those constructs that they self-reported as being noteworthy from this course. Whereas the immediate prior work deployed surveys with constructs that the author team determined best fit the context (Beem et al., 2023a), the author team sought here to allow the students' voice to inform the selection of constructs. Interviews were carried out with twenty-six participants who had participated in the course in the past. Questions asked in these interviews can be found in Table A3. Ten out of the twenty-six transcripts were determined to have enough substantive reflection to warrant further use. Participants who had had at least one year elapse since they took the course tended to provide deeper reflection in their responses. Twenty-four codes were derived from the ten transcripts, and these were grouped into the six categories stated above. These categories were expected to describe significant changes in skill and perception that students in this first-year engineering PBL course experience. Questions on Group Work Skills were obtained from the *"Development and Validation of group work skills questionnaire for higher education*" whilst questions on problem-solving skills were obtained from "*Development of Problem-Solving Confidence Questionnaire: Study of validation and reliability*" (Cumming et al., 2015; Gok, 2012). The other questions, such as Potential within Africa and Self, were drawn from insights gained in mock interviews run prior to the actual interviews.

The pre-survey was conducted in the first week of the course, and the retrospective pre-post survey in the last week of the course. The aim of this approach was to compare the same students' responses at different time points. That is to compare their pre-course score as they self-reported at both the beginning and end time points. Across both cohorts for the course, 82 students completed both surveys, representing approximately 87% of the whole class. Of these 82 students, 43 identified as male, and 39 identified as female. Students were also asked questions that captured whether they had previous design-build experience or not.

2.2 Quantitative analysis

Two sets of quantitative analyses were conducted on the survey data- one to determine the internal consistency of the surveys and one to determine any statistical significance between student responses at the various time points. Exploratory Factor Analysis (EFA) was used to determine whether the answers given by the students were correlated with others within the same category within each survey. A varimax rotation was conducted on each of the three sets of results (pre-survey, retrospective pre-test, and retrospective post-test). All factor loadings with an absolute value < 0.5 were ignored. For the sake of space, each question has a label corresponding to its section: GW (Group Work Skills), IP (Interpersonal Skills), PS (Problem-Solving Skills), PIA (Potential in Africa/Self), and OT (Other) questions.

Additionally, the statistical significance of the differences in the responses of the pre-survey and retrospective pre-test surveys, compared to those of the retrospective post-test surveys, was determined. To achieve this, paired two-tailed t-tests were used with a 5% threshold value. A Hedge's g-test was used

to determine the effect size of any statistically significant differences. The effect size was considered small if |g|>=0.2, medium if |g|>=0.5, and large if |g|>=0.8.

2.3 Qualitative analysis

From the quantitative analysis, it was seen that many participants rated themselves lower in the retrospective pre-test self-assessment than in the pre-survey. There was a need to understand the cause of the discrepancies between these two data points; hence, a qualitative approach was pursued. Participants who had the largest differences were identified. This was done by comparing their average scores within each transformation and comparing the difference between the two pre-course data points. Students whose scores were most distinct from each other were highlighted, emailed, and invited to an interview. Three people, 2 females and 1 male, volunteered and were then interviewed by the second author. The timeframe for these interviews was in February 2025, which was about ten months after the end of the course.

The interviews lasted roughly twenty minutes for each participant. The interview was semi-structured to elicit more insights into the differences between the pre-survey and retrospective pre-test self-assessments. The following four interview questions were used:

Interview Question 1: At the beginning of the class, you rated yourself a (score) out of 10 across the various categories. Can you reflect on what led to that self-assessment?
Interview Question 2: We observed an interesting trend across much of the class, which was that the pre-scores provided at the end of the course were relatively lower. You also fell into this trend, rating yourself around a (score) out of 10 in this. Can you reflect on what may have led to those differences for you?
Interview Question 3: Before starting the course, did you feel you had a clear understanding of its content in relation to the areas addressed in the survey? What do you think the course had to offer regarding the transformations?
Interview Question 4: In what ways, if any, did your perceptions of your skills change from the beginning to the end of the course?

The interviews took place online, with participants asked to select a meeting time of their convenience. The session was recorded on Zoom and transcribed using the Otter AI transcription tool. The interviewer then corrected discrepancies between the transcribed text and the audio recording.

Qualitative analysis was utilized to identify relationships and themes present in the data obtained through the interviews. This initial coding process utilized the descriptive and in vivo coding techniques. The in vivo analysis is best suited for this part of the study because it prioritizes the participant's voice. The descriptive code aims to answer the question "What is going on here?" (Saldaña, 2009, p. 70) and is also essential for assessing longitudinal participant change over time (Saldaña, 2003, 2008). In this study, the descriptive codes used were predetermined from the set of cognitive biases identified in Section 1.4. These codes were assigned to respondents' portions of the transcripts. The second level of coding was pattern coding, where similar and repetitive codes were categorized under overarching themes. Through pattern coding, this study was able to ascertain the major themes in the codes and patterns in the cause of discrepancies in the retrospective pre-test and pre-survey answers. Participant 1's transcript was coded to generate a codebook. The codebook consists of four columns: the code, the definition, the description, and an in vivo description of the code. During meetings with the principal investigator (fourth author), the elements in the codebook were critically assessed to ensure they captured the data from the participants. The codebook was then adopted and used to code the remaining transcripts.

The first cycle coding (structural coding) of the raw data developed the memocode, in which the following predetermined codes were assigned: Dunning-Kruger effect, overconfidence bias, anchoring bias, optimism bias, planning fallacy bias, Illusion of control bias, hindsight bias, self-serving bias, and impact bias. Pattern coding was utilized during the second-level coding process.

3 Results

3.1 Internal consistency of the surveys

The EFA carried out on the pre-survey answers yielded 9 factors that loaded haphazardly onto the survey questions. In each section of the survey, except Personal Evolution, there was one question that was not strongly loaded onto any factor. Despite the many factors, some categories loaded together on certain factors. Five of the seven questions in the Interpersonal Skills category loaded together, and five of the seven questions in the Personal Evolution category loaded onto one or more of two different factors. There are two factors that only have a strong negative correlation with one question each. Because of the volume of factors, it is difficult to assign a meaning to each factor.

The EFA carried out on the retrospective pre-test survey answers yielded 4 factors. This is shown in Table A4. The first factor loaded strongly onto all of the questions in the Group Work and Interpersonal Skills sections, as well as one question in the Problem-Solving section. This suggests that Group Work and Interpersonal Skills can fit together in one category of questions. The second factor loaded strongly onto every question in the Personal Evolution section and the first four questions in the Potential in Africa/Self section. This factor seems to depict individual self-assessment. The third factor loaded strongly onto one question in the Group Work section and all of the questions in the Potential in Africa/Self section. This suggests that the context of Africa makes it such that the first four questions in this section obtain information that is separate from self-assessment. The fourth factor loads strongly onto the first five questions in the Problem-Solving section, showing that the questions are correlated and give unique information. The sixth question in the Problem-Solving section is the only question without a strong loading onto any factor.

The EFA carried out on the post-retrospective survey answers yielded three factors. These overlapped across the various groupings. The first factor loaded strongly onto four of the seven questions in the Group Work section, one of the questions in the Interpersonal Skills section, five of the questions in the Problem-Solving section, eight of the nine questions in the Personal Evolution section, and all of the questions in the Potential in Africa/Self section. The second factor strongly loaded onto all of the questions in the Group Work Skills and Interpersonal Skills sections, along with one question in the Problem-Solving section and two questions in the Personal Evolution section. This further suggests that the Group Work and Interpersonal Skills questions extract similar information. The third factor only loaded strongly onto question six of the Problem-Solving section and question 2 of the Personal Evolution section.

Table 1 shows Cronbach's alpha scores for each set of survey responses. With each set of responses having a score of over 0.9, it can be said that these surveys are reliable.

Table 1: Cronbach's alpha scores for survey questions

Survey	Cronbach Alpha	Number of Questions
Pre-Survey	0.941	40
Retrospective Pre-test	0.986	36
Retrospective Post-test	0.988	36

3.2 Comparison of the two surveys

There were no statistically significant differences between the answers given on the pre-survey and the retrospective post-test survey. If one were to look solely at that dataset, one may determine that this class had no impact on the students. However, based on the comparison of the data within the retrospective survey (retrospective pre-test and the retrospective post-test), statistically significant increases were achieved across the board. As shown in Table 2, statistically significant increases were seen in all five categories. This happened with a small effect size on Group Work Skills, Interpersonal Skills, and Problem-Solving Skills. This happened with a medium effect size on how much the students believed that they had grown and the potential of engineering to produce positive change within Africa.

Table 2: p-values of differences in results between retrospective pre-test and retrospective post-test data

Category	P-value	Hedge's g value	Effect Size
Group Work Skills	6.34 E-10***	0.454	Small
Interpersonal Skills	5.07 E-06***	0.201	Small
Problem-Solving Skills	1.90 E-11***	0.426	Small
Potential in Africa/Self	4.83 E-17***	0.572	Medium
Other	1.13 E-16***	0.582	Medium

**p<= 0.05*
***p<= 0.01*
****p<=0.001*

These trends are largely mirrored in the disaggregated data. Across both genders, both cohorts, and students with/without previous design-build experience, statistically significant increases with small to medium effect sizes were measured. Exceptions to this include the Interpersonal Skills category, which saw no effect size for the following disaggregations: Females, Cohort B, and students with previous experience. Students with previous experience only showed an increase of small effect size in Potential in

Africa/Self. And for these two disaggregations- Females and students with no previous experience- there was an increase of medium effect size in Problem-Solving Skills.

3.3 Qualitative Results

In the interviews, the participants reported significant gains from the course. For example, Participant 2 noted substantial improvements in her skills. She stated, "*We had experience with SolidWorks. We learned about the design cycle and programming in C*." The tasks assigned in class, which initially seemed out of scope, helped her develop an anticipation for future tasks and laid a strong foundation for her engineering journey. She said, "*Building a rocket is obviously not something that first-year students typically do. So, at the beginning of the semester, even though we didn't know the full scope of what we were doing, we had an idea.*"

The gains they articulated aligned with the changes reported on the retrospective pre-post survey. Coding of the interview data revealed nine biases that contributed to discrepancies between the two datasets in this study. These nine biases were coalesced into three overarching categories: Self-Assessment Bias, Memory/Recall Bias, and Influence/Response Bias.

Self-Assessment Bias: From the interview data, the Dunning-Kruger effect was prevalently exhibited by participants as the cause of the discrepancy in their pre-survey and retrospective pre-test assessments. All three participants said they rated themselves highly in the pre-survey due to their experiences in high school. Participant 2 specifically mentioned her high school Physics experience and the fact that some concepts at the initial stages of the course were related to that; hence, she rated herself highly. The high school experience, coupled with initial experiences in the first few days of taking the class, also constituted an anchor bias- the situation where participants rely on previous experiences (the anchor) to assess present conditions. She went on to say, "*I guess [the high score in the pre-survey*] *was based on what I felt I had experienced, not exactly proficiency, but I mean, I've encountered this before*." Participant 3 also rated himself highly in the retrospective pre-test survey because, before coming to the university, he had always known himself to be a good student. Some participants also exhibited an overconfidence bias. Participant 2 was overly confident in her abilities as a "*fast learner*," believing this quality would help her overcome any potentially challenging topics that might arise in the course, such as programming. With this initial belief, she rated herself highly (an average of 8.34 out of 10) in the pre-survey. After realizing that some more complex skills are harder to acquire, she then rated herself relatively lower (an average of 7.13 out of 10) in the retrospective pre-test survey.

Memory/Recall Bias: Participants also exhibited hindsight bias- they anticipated the course to be easy before taking it and realized there was more to it than they initially envisaged. The anticipation of ease led to relatively higher ratings in the pre-survey, and they readjusted the score in their retrospective pre-test survey. Participant 2 mentioned, "*I wouldn't say I entirely struggled completely in Intro* [to Engineering]*, but it wasn't as easy as it looked like at the beginning of the semester.*" As part of the course, students were exposed to coding in C++. Some participants were excited about the prospects of learning how to code; however, they were unaware of the challenges that could arise from this new endeavor. The initial excitement led to high ratings in the pre-survey. Participant 1 mentioned, "*I realized that [programming] wasn't one of my strong suits.*" She conceded to having overestimated herself going into the course. "*I would say I overestimated how smooth the course would be for me. Yes, I overestimated that.*"

Influence/Response Bias: There were elements of careless responses in some of Participant 3's responses, which adversely affected his overall average. Even though he could not justify his self-confidence or how

well he worked in a team, he went ahead to complete the survey with some responses, and conceded that "*I don't know why I rated myself lower*." This concession may be due to his forgetting his scores in both surveys or his desire to express socially desirable qualities, which may not have entirely represented his reality. He may have felt he needed to find a way of reporting a transformation after taking the course, even if it was not the case. Such response biases can present inaccurate self-ratings and end up altering the quality of the data. The length of the survey questions may have also contributed to the careless responses in Participant 3's responses. This study had 44 and 81 Likert-scale and short-phrase questions in the pre-survey and retrospective survey, respectively.

These results are summarized in Figure 1, which depicts the three overarching biases elicited through the interviews and their components.

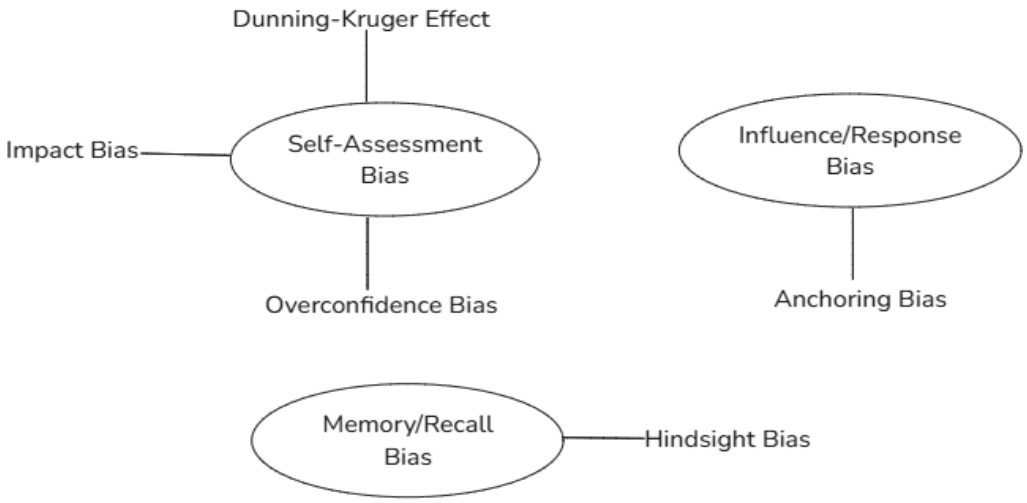

Figure 1: The overarching biases and their components.

Discussion and Conclusion

4.1 Limitations

The qualitative aspect of the research was limited by the sample size and the time that had elapsed between the end of the course and the interviews. Although 20 potential participants were identified through the steps outlined in section 2.3, only 3 signed up to participate, despite multiple email reminders. While this number is generally adequate for qualitative studies (Robinson, 2014), a larger sample might have generated more insights and enriched the findings. To boost sample size in future studies, the authors will consider offering rewards such as phone credit, shopping coupons, or souvenirs to motivate greater participation. Additionally, the course concluded in April 2024, while the qualitative phase did not begin until February 2025, nearly a year later. The demands of their current courses might have overshadowed their memories of their previous courses, making it difficult for them to accurately recall details of their experiences.

4.2 Future Work

This mixed-method study with first-year engineering students in Ghana has shed insight into relevant survey questions and survey types for their context. EFA on the newly developed set of survey questions

was shown to be internally consistent. Factor loadings on the retrospective pre-test were clean and can be used to identify improvements that can be effected in the next deployment.

The factor loadings seen in the retrospective post-test differed somewhat. This could be due to the fact that there are a large number of students answering “10” on the survey questions, leading to some form of ceiling effect. More investigation should be done to identify any actions to be taken to reduce the impact that this effect might have on future surveys.

4.3 Conclusion

All five skill/perception categories showed statistically significant increases with small to medium effect sizes in aggregate. This indicates that each of these does indeed align with what students experience and grow through this course. Interpersonal Skills is perhaps the weakest of these, as it showed no effect size for select disaggregations. Potential in Africa/Self showed one of the greatest changes and is therefore worthy of continuing to dig deeper to uncover the exact causes and mechanisms for this, so that other African universities can similarly cultivate this in their student population. The increase in this category had, however, a lesser effect for students who had some previous design-build experience.

Although the response-shift bias was expected to be a contributor to differences between the two survey types deployed here, it appeared to play a lesser role than the self-assessment bias in this study. The Dunning-Kruger effect and related biases in which the respondents overestimated their ability at the beginning came out the most strongly. The anchoring bias and hindsight bias also emerged as secondary biases at play. Given the small sample size, these insights should be interpreted as pointers to what can be further investigated in the future.

These results suggest that the retrospective pre-post survey may be superior to the traditional pre-post approach in this context. This was seen in a few ways: 1) Although both survey methods were reliable, the retrospective surveys had a slightly higher Cronbach's alpha. 2) Data from the retrospective surveys loaded cleanly onto a few factors, compared to the many factors produced by the pre-survey. 3) The retrospective surveys showed statistically significant increases in students’ perception of their abilities upon the completion of the course, while the pre-survey did not. This, combined with the fact that students in interviews stated that the course did have some transformative effects on them, indicates that the retrospective survey approach is more effective in capturing this transformation. Although likely superior, the retrospective survey may not be entirely exempt from biases, such as the ceiling bias. Perhaps for this or similar target populations, researchers should always consider including qualitative work to supplement surveys so as to cross-check any insights gained. Additionally, similar studies should be conducted in other African institutions to ascertain the extent to which the experiences captured in this study translate more broadly.

5 References

Ampomah, C. O. A., & Beem, H. (2024). Full Paper: Continued exploration of the relevance of self-efficacy, self-determination, and agency in describing the first-year African engineering student's experience. *15th Annual First-Year Engineering Experience Conference (FYEE)*, Boston, Massachusetts. https://doi.org/10.18260/1-2--48592

Baer, R. A., Ballenger, J., Berry, D. T. R., & Wetter, M. W. (1997). Detection of random responding on the MMPI-A. Journal of Personality Assessment, 68(1), 139-151. https://doi:10.1207/s15327752jpa6801_11

Beem, H. R. (2021a). Exploring the role of project-based learning in building self-efficacy in first-year African engineering students. In *2021, ASEE Virtual Annual Conference Content Access*.

Beem, H.R. (2021b). The role of machine shop sessions in building confidence amongst first-year engineering students in Ghana. *IEEE Global Engineering Education Conference, EDUCON*, *2021-April*(April), 1469–1474. https://doi.org/10.1109/EDUCON46332.2021.9454012

Beem, H. (2022). Exploring the role of the physical environment in building self-efficacy in first-year African engineering students. *ASEE Annual Conference and Exposition, Conference Proceedings*.

Beem, H., Ampomah, C. O. A., Takyi, J. P. K., & Adomdza, G. (2023a). A preliminary exploration of the relevance of self-efficacy, self-determination, and agency in describing the first-year African engineering students' experience. In *2023 ASEE Annual Conference & Exposition*.

Beem, H. R., Ampomah, C., Takyi, J., & Adomdza, G. K. (2023b). Development of an Online Project-Based Learning Design Course for African First-Year Students and Its Impact on Self-Efficacy Levels. *IEEE Transactions on Education*, *66*(5), 410–420. https://doi.org/10.1109/TE.2023.3312968

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI2 random responding indices: Validation using a self-report methodology. Psychological Assessment, 4(3), 340-345. https://doi:10.1037/1040-3590.4.3.340

Christensen, T. C., Wood, J. V., & Barrett, L. F. (2003). Remembering everyday experience through the prism of self-esteem. Personality and Social Psychology Bulletin, 29(1), 51–62. https://doi.org/10.1177/0146 167202238371

Cumming, J., Woodcock, C., Cooley, S. J., Holland, M. J. G., & Burns, V. E. (2015). Development and validation of the groupwork skills questionnaire (GSQ) for higher education. *Assessment and Evaluation in Higher Education*, *40*(7), 988–1001. https://doi.org/10.1080/02602938.2014.957642

Davis, G. C., Baral, R., Strayer, T., & Serrano, E. L. (2018). Using pre-and post-survey instruments in interventions: Determining the random response benchmark and its implications for measuring effectiveness. *Public Health Nutrition*, *21*(6), 1043-1047.

Drennan J & Hyde A (2008) Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme. Assessment and Evaluation in Higher Education,33 (6):699-709

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology* (Vol. 44, pp. 247-296). Academic Press.

Evans, C. M. (2022). What About Pre-and Post-Motivation Survey Scores? How Might They Support My Conclusions? *Reading Motivation: A Guide to Understanding and Supporting Children's Willingness to Read*, 91.

Ferguson, A., Gilkey, B., Kern, D., & Jasmine, D. (2012). An education initiative concerning lead in an Arkansas community: results from pre-and post-surveys. *Journal of Community Health*, *37*, 599-609.

Forsyth, D. R.(2008). Self-Serving Bias. *International Encyclopedia of the Social Sciences*. Edited by William A. Darity. 2nd ed. Vol. 7. Detroit: Macmillan Reference USA.

Gok, T. (2012). Development of Problem Solving Confidence Questionnaire: Study of validation and reliability. *Am. J. Phys. Educ*, *6*(1). http://www.lajpe.org

Grimes, A., Medway, D., Foos, A., & Goatman, A. (2015). Impact bias in student evaluations of higher education. *Studies in Higher Education*, *42*(6), 945–962. https://doi.org/10.1080/03075079.2015.1071345

Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. American Journal of Evaluation, 26(4), 501–517.

Howard, G.S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4, 93-106.

Howard, G. S., Ralph, K. F., Gulanick, N. A., Maxwell, S. E., & Gerber, S. K. (1979). Internal validity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. Applied Psychological Measurement, 3(1), 1–23.

Hufford, J. R. (2010). What are they learning? Pre-and post-assessment surveys for LIBR 1100, Introduction to library research. *College & Research Libraries*, *71*(2), 139-158.

Johnson, D. D. P., & Fowler, J. H. (2011). The evolution of overconfidence. Nature, 477(7364), 317–320. https://doi.org/10.1038/nature10384

Lamb, T. (2005). The retrospective pretest: An imperfect but useful tool. *The evaluation exchange*, *11*(2), 18.

Kowalski, M. J. (2023). Measuring changes with traditional and retrospective pre-posttest self-report surveys for a brief intervention program. *Evaluation and Program Planning*, *99*, 102323.

Lay, J. C., Gerstorf, D., Scott, S. B., Pauly, T., & Hoppmann, C. A. (2017). Neuroticism and extraversion magnify discrepancies between retrospective and concurrent affect reports. Journal of Personality, 85(6), 817–829. https://doi.org/10.1111/jopy.12290

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. Psychological Methods, 17, 437-455.

Mill, A., Realo, A., & Allik, J. (2015). Retrospective ratings of emotions: the effects of age, daily tiredness, and personality. Frontiers in Psychology, 6, 2020. https://doi.org/10.3389/fpsyg.2015.02020

Moore, D., & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation*, *30*(2), 189-202.

Nimon, K. (2014). Explaining differences between retrospective and traditional pretest self-assessments: competing theories and empirical evidence. International Journal of Research & Method in Education, 37(3), 256–269. https://doi.org/10.1080/1743727X.2013.820644

Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, 21(4), 331–339. https://doi.org/10.1177/0146167295214003

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (pp. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. New Directions for Institutional Research, 121, 63–73.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative research in psychology*, *11*(1), 25-41.

Saldaña, J. (2009). *The coding manual for qualitative researchers.* Sage Publications Ltd

Schwartz, C. E., & Sprangers, M. A. (2010). Guidelines for improving the stringency of response shift research using the thentest. *Quality of Life Research*, *19*, 455-464.

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 349–363). Oxford University Press. https://doi.org/10.1093/med:psych/9780199356942.003.0024

Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. Psychological Science, 14(5), 520–524. https://doi .org/10.1111/1467-9280.03455

Zurbriggen, C. L. A., Jendryczko, D., & Nussbeck, F. W. (2021). Rosy or blue? Change in recall bias of students' affective experiences during early adolescence. *Emotion, 21*(8), 1637–1649. https://doi.org/10.1037/emo0001031

Appendix

Table A1: Influence/response biases and their characteristics by Wetzel et al. (2016)

Influence/Response Bias	Characteristics
Acquiescence response style	Preference for categories stating agreement (e.g., agree, strongly agree).
Disacquiescence response style	Preference for categories stating disagreement (e.g., disagree, strongly disagree).
Careless responding	Inattentive responding.
Extreme response style	Preference for extreme categories (e.g., strongly disagree, strongly agree).
Midpoint response style	Preference for the midpoint of a rating scale (e.g., neutral).
Socially desirable responding	Tendency to describe oneself positively to reflect social norms and rules.

Table A2: List of survey questions asked, grouped by Transformation type

TRANSFORMATION	QUESTIONS
Group Work Skills	I remind the group how important it is to stick to schedules.
	I construct strategies from ideas that have been raised.
	I clearly define the roles of each group member.
	I move the group's idea forward towards a strategy.
	I evaluate how well the group is progressing toward the agreed goals.
	I provide constructive feedback on the group's progress towards agreed goals.
	I dominate in group discussions.
	I encourage everyone to set aside their personal agendas or their personal interest.
Interpersonal Skills	I provide emotional support to my group members.
	I am sensitive to the feelings of other people.
	I show that I care about my group members.
	I can be open and supportive when communicating with others.
	I can be there for other group members when they need me.
	I encourage the group to set goals that are challenging and attainable.
	I believe that group members are capable of completing their share of the work.

Problem-Solving Skills	I am sure that I can solve a problem.
	I try too hard when I cannot solve the problem.
	I am self-confident in problem-solving.
	I do my best to be successful in problem-solving.
	I struggle with a problem until I find the correct answer.
	Preconceptions prevent me from solving problems.
	I am interested in problem-solving.
	I enjoy solving a problem.
	I like to solve a problem.
	I do my best to solve the problem, no matter how difficult the problem is.
Potential in Africa/Self	I see myself as capable of solving problems around me.
	I am capable of designing and building solutions to problems I identify around me.
	I see myself as a local innovator.
	I see myself as capable of creating technologies that will improve lives.
	I see the possibility of technological advancement in Africa.
	I see Africa as a place where technology and innovation can be produced.
	I believe Africa's technical challenges can be addressed by local engineers.
	I see myself being part of the engineering workforce that leads Africa's development.
Other	I see engineering to be difficult.
	I have the ability to recover quickly from difficulties when working on an engineering project.
	I can continue working on an engineering project even though it is difficult.
	I have the ability to commit myself to my engineering project.
	I am sure of my ability to work on an engineering project.
	I want to know or learn more about engineering.
	I take charge of my learning.

Table A3: List of interview questions prior to survey

Interview Questions
1. As a first-year African student, can you describe your expectations towards enrolling in a course that gives you practical experience?

2. How important was it for you to get good grades in the course, and why?
3. Can you describe the usefulness/helpfulness of the course to you in the first year?
4. Can you describe your level of interest in the course?
5. In what ways did you adjust to any challenge you may have faced in the course?
6. What did you learn about yourself as you took the course?
7. How did the course help you make real-world connections?
8. What did you learn about yourself by working with other people?
9. How did the course affect your mindset toward engineering?
10. In what ways did the course affect how you see yourself? (eg. Confidence, competence)
11. What activities are you currently engaged in that were inspired by the course?
12. How did the course affect your ability to make decisions?
13. How did your experiences result in your developing an interest in the course?

Table A4: Factor loadings of retrospective pre-test survey questions

	Component			
Question	**1**	**2**	**3**	**4**
GW01	.691			
GW02	.655			
GW03	.553		.583	
GW04	.613			
GW05	.725			
GW06	.679			
GW07	.705			
IP01	.778			
IP02	.819			
IP03	.836			
IP04	.821			
IP05	.773			
IP06	.730			
PS01				.509
PS02				.620
PS03				.658
PS04				.572
PS05	.526			.571
PS06				
OT01		.574		
OT02		.772		
OT03		.711		
OT04		.737		
OT05		.666		
OT06		.755		
OT07		.604		
OT08		.685		
OT09		.640		
OT10		.624		
PIA01		.600	.568	
PIA02		.592	.512	
PIA03		.550	.622	

PIA04	.509	.664
PIA05		.771
PIA06		.700
PIA07		.810