Student Learning in the Natural Sciences

A Correlation Assessment

Phillip Vargas
Harold Washington College Assessment Committee
Harold Washington College
Chicago, IL
pvargas21@ccc.edu

***Abstract*—The Harold Washington College Assessment Committee administered an assessment in the natural sciences. This assessment was a newly developed tool to measure student learning, attitudes about the sciences, and how they correlate with academic history. This tool was deployed entirely electronically and cross-referenced with the City Colleges of Chicago Openbook database.**

The results of this assessment showed that taking classes in the natural sciences both improved student performance on the assessment tool and positively affected their attitudes about the natural sciences. It also illustrated the complexity of attempting to quantify these gains in an academic structure that has multiple paths and entrance points.

I. Introduction

In November of 2015, the Harold Washington College Assessment Committee (HWCAC) administered an assessment in the college's general education category of the natural sciences, which span both the physical and biological sciences. This was the second time the assessment committee assessed learning in the physical sciences and the first time in the biological sciences. The physical sciences were last assessed in 2007 with the Epistemological Beliefs Assessment for Physical Science (EBAPS) tool. Debate regarding the trade-offs between administering EBAPS again and obtaining comparative data versus developing a new tool and building on seven years of assessment experience, led to the decision to proceed with a developing a new tool.

While the EBAPS is an excellent tool for probing student views of knowledge and learning in the physical sciences, it was decided that ascertaining how much learning is occurring better aligns with the mission of the assessment committee. Therefore, the goal of the new assessment was to more accurately measure learning in the natural sciences and attitudinal shifts based on students' academic history. This information could better allow the committee to make recommendations to improve curriculum design and direct support services.

II. Methodology

A. General Education Goal and Departmental Structure

The general education goal for the natural sciences is to understand the major principles of the natural sciences and the application of the scientific method to biological, physical, and environmental systems. To meet this goal, the following student learning objectives have been codified buy the HWCAC with support from the Department of Physical Sciences and the Department of Biological Sciences:

1. Formulate reasonable explanations of natural phenomena based on thorough observations.
2. Interpret and articulate scientific results that are presented in verbal, graphic and/or tabular form.
3. Critically evaluate scientific resources and scientific claims presented in the media
4. Apply steps of the scientific method to solve problems.

The natural sciences in Harold Washington College encompass two departments: The Department of Biological Sciences which offer classes that fulfill the Illinois Articulation Initiative General Education Core Curriculum (IAI GECC) requirements in the life sciences and the Department of Physical Sciences which satisfy the IAI GECC requirements in the physical sciences. Both of these departments offer more individualized academic disciplines such as physics or microbiology. An organization of these disciplines can be seen in figure 1.

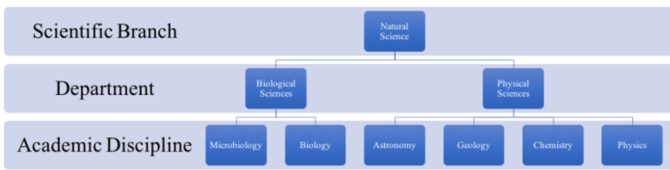

Fig. 1 Academic disciplines in the Natural Sciences at Harold Washington College

At the course level, the course subjects may correspond to either the academic discipline such as Physics 235 or to the department such as Physical Science 112 which is a topics course that covers multiple academic disciplines. The Department of Physical Sciences major courses are confined to the two academic disciplines Physics and Chemistry. While the other academic disciplines as well as the general purpose course subject "Physical Science" are indented for non-major students. In the Biological Sciences, the course subject "Biology" contains both major and non-major classes. As both Physical Science and Biology can define academic disciplines or course subjects this report will write out the qualifier as well. It should be noted that the difficulty of attempting to define these terms to be congruent with scientific nomenclature, IAI GECC terminology, and the City Colleges of Chicago's Academic Catalog could make navigating, enrolling, and registering for these courses challenging. While attempting to restructure and reword the academic catalog may not be practical or possible, the department could design pamphlets to better explain the courses within their department and their intended audiences. This could support students and advisors during course registration.

B. Tool Development

To determine the degree of learning in the natural sciences, a twenty question, multiple choice bank was written: ten questions in the physical sciences and ten questions in the natural sciences. The number of questions chosen was an attempt at optimizing between testing fatigue and obtaining a sufficiently large sample to acquire valid and reliable data. A heuristic approach was taken in estimating question duration and attempting to keep the survey under thirty minutes.

When reviewing the student learning outcomes for the natural sciences alongside the structure of the courses offered in the Departments of Biological and Physical Sciences several immediate challenges became apparent. These departments offer a large breadth of topics with each course providing depth in its respective discipline. Thus, overlap across the courses was not immediately apparent. Additionally, while the student learning objectives for the natural sciences are process based, most of the student learning objectives at the course level are content based. Developing process based questions that mirror the general education student learning outcomes and encompass the content based student learning outcomes was determined to be too large of an undertaking to be accomplished in the proposed timeframe. Additionally, finding the intersection of these disciplines aside from the definition of science, which in itself is a nebulous concept, was also determined to not be the best course of action. Ultimately, fundamental concepts expressed in course student learning outcomes from the primary disciplines were chosen for the content questions.

The questions from the physical science portion covered Geology, Meteorology, Astronomy, Chemistry, Physics, and the scientific method. These questions were adapted from common assessment tools in their respective disciplines: the Astronomy Diagnostic Tool 2.0, Force Concept Inventory, Energy Concept Inventory, and Survey of Public Attitudes Toward Understanding of Science and Technology. The questions for the biological science portion covered life hierarchy, flow and storage of heredity information, metabolism, cell division/mutation, and genetics. These questions were designed in-house by faculty members of the Department of Biological Sciences.

In addition to content questions, affective questions were included in the tool. These questions were taken from the Colorado Learning Attitudes about Science Survey (CLASS). This tool builds on the Maryland Physics Expectation Survey (MPEX), Epistemological Beliefs Assessment for the Physical Sciences (EBAPS), Views of Nature of Science (VNOS), and Views About Science Survey (VASS). The CLASS tool has 42 questions, but only 27 of them have been validated and linked to affective categories. 12 of these questions were removed due to having high discipline specialization, high correlations with other questions, and low weights in principle component analyses. The remaining questions provided a bank of 15 Likert-scale questions for the affective portion.

With the content questions, affective questions, and request for consent and student identification number the tool had a total of 37 questions. All demographic information and academic history would be obtained from student identification and the Openbook database. Based on previous tools the questions requiring higher order cogitative thinking were placed before the affective questions to improve validity.

C. Data Acquisition and Processing

A complete electronic data acquisition process was implemented for this assessment. Google forms was used to acquire student identification numbers and responses to the assessment. The goal of this was to reduce the necessity of requesting faculty to volunteer their classes and to get a larger response rate. Additionally, this was the first time students would be asked to provide their student identification numbers for the use of cross-referencing with the Openbook database. The goal of this was to improve the validity of their academic history as it will no longer be self-reported and to reduce the testing time.

Although faculty members were not required to volunteer their classes, many encouraged students to participate in the assessment. Many faculty members offered extra credit, posted reminders in their learning management systems, and allocated instructional time for them to complete the assessment. Without these efforts and multiple reminders from the assessment committee a representative sample would not have been achieved.

While acquisition goals were met, extracting the information from Openbook required significant post-processing. Of the students participating in the assessment only 45.9% of classes associated with them had a letter grade attached to those classes. While the outcome in the class could be determined from other fields in the database, the classes needed to be downloaded for processing. To aid in the processing a new structured database was developed in Python and C to format and clean the data and allow for statistical and analytic processes beyond what is standard in the Openbook API. As this data can be traced back to individual students, privacy was ensured by anonymizing the data and storing it on an AES 256-bit encrypted disk volume. These techniques exceed all state and federal guidelines for "data at rest" compliance.

III. Validation

A. Pilot Studies

The Natural Science Assessment Tool was first piloted to the Harold Washington College Assessment Committee-at-large in Spring 2015. The sample for this study consisted of 18 faculty members and the goal was focused on improving readability. No significant changes to the tool came from the results of this study, but some questions were rephrased to improve clarity. A second pilot study was conducted on the student population in the summer of 2015. This study involved 103 students who were enrolled in physical sciences and astronomy courses. The primary goal of this pilot study was to test the efficacy of using student IDs and the Openbook software platform. No significant obstacles were encountered, and the assessment was cleared for administration for fall 2015.

B. Student Sample

The student populations enrolled in credit courses for Harold Washington College and the City Colleges of Chicago in Fall 2015 were 9,116 and 36,977 respectively. A total of 1,050 respondents completed the Natural Science assessment. Of these respondents 982 (93.5%) had unique student identification numbers and 956 (91.0%) of these unique identification numbers matched student identification numbers matched students in the Openbook database. With 10.5% of the HWC's population completing the survey, a margin of error of 2.7% at 95% confidence errors was obtained. As methods for determining margins of error vary, for clarity the calculation in this analysis is

$$e = \sqrt{\frac{z^2 p(1-p)}{n\left(1-{}^{n}/_{N}\right)}}$$

where e is margin of error, z is z-score, p is percentage value, n is sample size, and N is population size. The percentage value used was the most conservative 0.5 value as no *a priori* knowledge was assumed.

Demographic categories were obtained from Openbook to confirm that the sample was representative of Harold Washington Colleges population in terms of gender, race, and age. Confirming the sample was representative in these categories was completed by performing hypothesis testing on the sample and population frequencies. Gender (Figure 1) and ethnicity (Figure 2) were tested with the Wilcox Rank-Sum Test and Age (Figure 3) with the Kolmogorov-Smirnov Test. None of these tests found a statistically significant difference between the sample and population in this assessment. Based on these results it was determined that stratification methods would not be necessary and the sample could be used as acquired in its entirety.

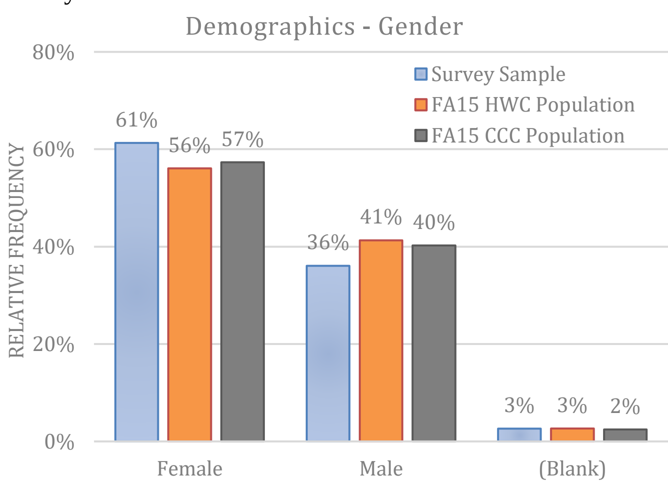

Fig. 2 Relative frequency of gender from the survey participants and college(s) population. All data was self-reported by students and collected through Openbook.

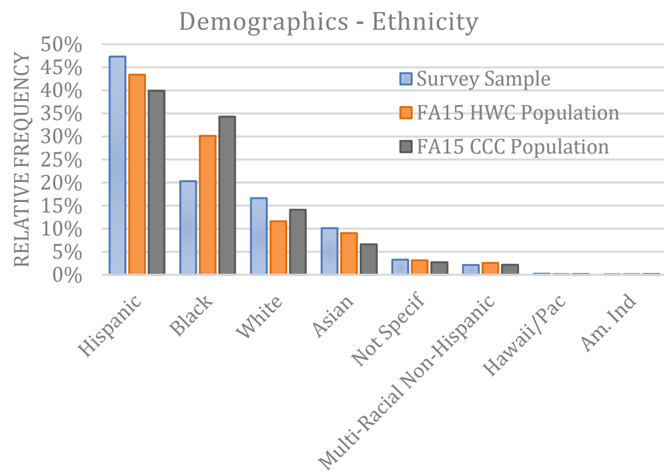

Fig. 2 Relative frequency of ethnicity from the survey participants and college(s) population. All data was self-reported by students and collected through Openbook.

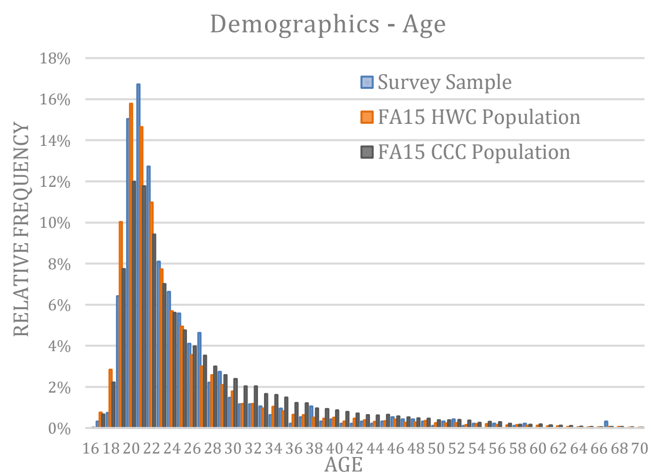

Fig. 3 Relative frequency of age from the survey participants and college(s) population. All data was self-reported by students and collected through Openbook.

C. Internal Reliability

The Cronbach α test was used to determine the lower bound on the reliability of content portion of the tool, and a value of 0.65 was calculated. Additionally, a point-biserial correlation coefficient test was conducted on each of the content question with representative values found in Table 1.

Table 1 Point-Biserial Results

Pt. Biserial	**Physical Science**	**Biological Science**
Mean	0.46	0.42
Minimum	0.36	0.10
Maximum	0.52	0.58

These results show there is room for refinement with this tool. The "border line" Cronbach α value and varied point-biserials illustrate that the tool could benefit from calibration in difficulty of the questions as well as from having more questions. This was a concern in the development stages as the tool would have to span such a large breadth of disciplines.

IV. Findings

A. Performance Distributions Department Level

The tool was developed to be challenging and have a mean score around 50%. The aim of this was to maximize the spread of scores and minimize ceiling and floor effects. This score on the content portion is not calibrated to typical letter grades and should not be interpreted as such. While some anecdotal evidence and percentile charts were used to obtain a mean score of 50%, a robust calibration technique was not implemented. Modifications were planned after the summer 2015 pilot, but with a mean score falling within one standard deviation of 50%, additional fine tuning would not have been possible. The performance for the distribution on the college-wide assessment can be seen in figure 4.

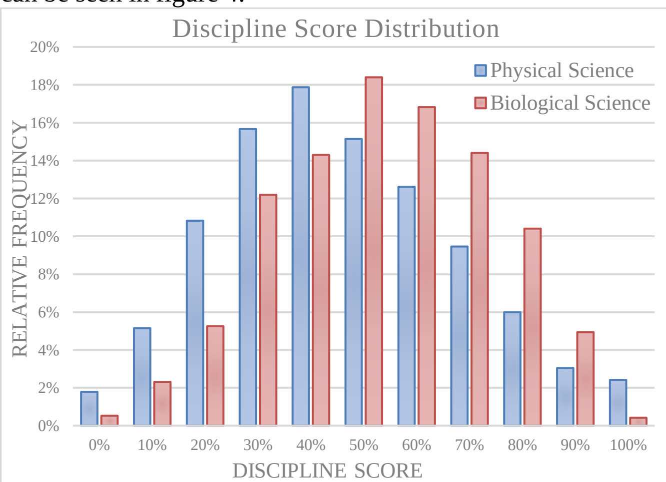

Fig. 4 Performance distribution on the content portion of the assessment separated by discipline.

The score distributions were statistically different than an ideal normal distribution based on a D'Agosostino's K^2 Normality Test. However, this statistical difference is likely due to the relatively large sample size and for the purposes of this analysis is being used to solely determine the hypothesis testing techniques. The mathematical moments which can be found in Table 2 of the two distributions were determined to be close enough to the ideal that parametric testing techniques could be implemented.

Table 2 Performance Distribution Moments

Moment	Physical Science	Biological Science
Mean	46.39%	53.25%
Variance	4.96%	4.05%
Skew	0.28	-0.09
Kurtosis	-0.39	-0.59

These distributions indicated that this tool would be a good candidate for the non-major physical and biological science topics courses that are currently not being assessed at the department level. If the relevant sections of the tool are used for these classes, there will be considerable comparative data available with this report.

B. Performance Distributions – Academic Discipline Level

Students performed worst on the physics (38%) and astronomy (39%) questions and best on the general science (51%) and meteorology (49%) questions in the physical science portion. In the biological sciences portion students performed best on levels in life hierarchy (62%) and worst in cell division/mutation (44%). As only two questions were chosen for each of these disciplines, it is not possible to determine if these scores were more influenced by the difficulty of the topic, exposure to the topic, or difficulty in the specific question posed. These weights could be estimated by follow-up assessments at the department level.

There is also the possibility that these courses are being taught without emphasis on application in regards to physics and chemistry, or without fundamental principles in regards to geology, astronomy, and meteorology. As the A.A, A.G.S, and A.F.A all only require one semester of a physical science and biological science course this could be potentially the last formal exposure of these concepts to many of our students.

When examining the performance on content questions within a specific discipline's course sequence mixed results were found. For example, when examining performance on the chemistry content questions as a function of the highest chemistry course taken, the performance was not monotonically increasing as seen in Table 3.

Table 3 Chemistry Content Performance

Highest Chemistry Course	Performance	Standard Deviation	n
Organic Chemistry II	75.00%	25.00%	10
Organic Chemistry I	64.81%	29.86%	27
General Chemistry II	72.22%	29.91%	54
General Chemistry I	69.51%	30.33%	123
Basic Chemistry	68.49%	31.00%	119

However, these performances are significantly higher than for students that have never taken a chemistry course which is 52.50% with a standard deviation of 35.78% and an *n* of 779. The roughly 17% increase in performance suggests that by taking these courses students develop a better understanding of fundamental chemistry concepts. However, the chemistry content questions posed in the assessment tool were designed to be "moderately challenging" for any student exposed to chemistry principles in the Department of Physical Science, and "easy" for students taking chemistry-major courses. The results may suggest that instructors assume these fundamental concepts are already known in higher level courses, and therefore may not further develop or reinforce them in subsequent courses. It may also suggest that the basic content questions are not capturing the sophistication learned in subsequent courses.

C. Correlation between Course History and Assessment Performance

A correlation analysis was performed between courses taken in each of the 87 disciplines offered at the City Colleges of Chicago and the performance on the natural sciences assessment (Figure 5). While all of these disciplines cannot be illustrated for layout reasons, four of these disciplines have positive correlations that are statistically significant based on an ANOVA with a Tukey-Kramer *post hoc* analysis. These disciplines include Chemistry, Biology, Physics, and Inter-disciplinary (INTDSP 299 is the course designation in the STEM Scholars program.) While this result was confirmation that learning in the natural science is occurring in some of the STEM disciplines, it should be noted that not all of the disciplines were present. Additionally, multiple disciplines have negative correlations associated with them. Two of these include mathematics and physical science.

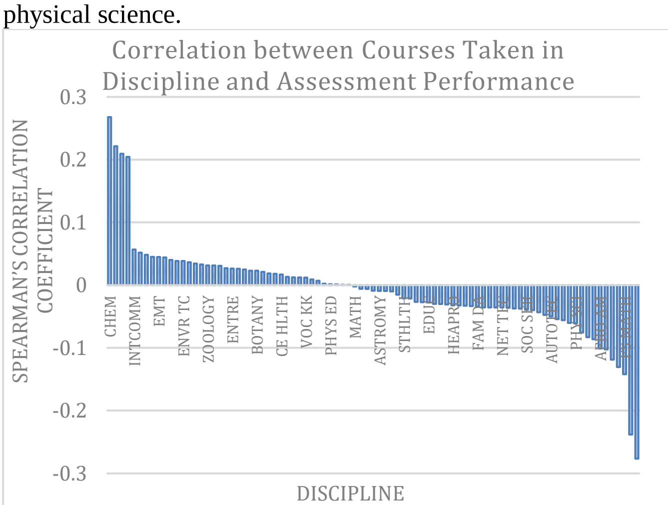

Fig. 5 Spearman's Correlation Coefficent for the 87 distinct disciplines offered at the City Colleges of Chicago

These results state that a student who completes a physical science course performs worse (-0.06 Spearman's correlation coefficient) on this assessment compared to a student who doesn't complete a physical science course. At face value this would appear as though not only are students not learning in these courses, but they are performing worse. However, this interpretation does not take into account the complexity of the academic tracks at Harold Washington College and the City Colleges of Chicago more generally. Students taking courses in STEM programs do not take courses within the Physical Science discipline because those courses are intended for non-majors. STEM students' high performance on this exam (Figure 6) skews these results. When compensating for this by only comparing non-STEM students, these coefficients increase by 0.1 in physical science and astronomy, thus showing that learning in the natural sciences is occurring from taking these classes. Controlling for these variables was not possible during the 2007 natural science assessment and most likely led to skewed results when comparing results based on the number of courses taken at Harold Washington College to other institutions.

Additionally, the most negative correlations were associated with developmental classes: English 100 (-0.30), Reading 125 (-0.28), Math 98 (-0.22), Math 99 (-0.22), and English 197 (-0.19). These results were similar to findings the HWCAC's Effective Writing report. However, while those results were somewhat expected for a writing tool, it was not predicted to play as a large of a role in the natural sciences. These results are suggesting that students entering the City Colleges of Chicago at the developmental level in mathematics or English do not just underperform in assessment tools for those disciplines, but in other general education disciplines as well.

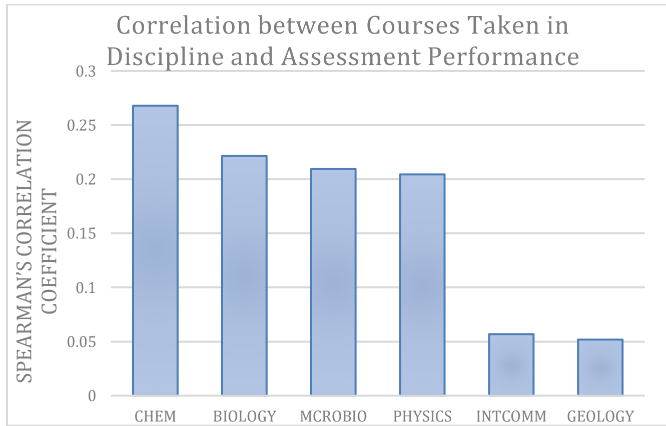

Fig. 6 Spearman's Corrleation Coefficent for highest six disciplines offered at the City College of Chicago

D. Attitudinal Shifts

The affective portion of the assessment focused on eight categories: Personal Interest, Real World Connection, Problem Solving General, Problem Solving Confidence, Problem Solving Sophistication, Sense Making/Effort, Conceptual Understanding, and Applied Conceptual Understanding. While measuring these shifts is most effectively accomplished in pre- and post-surveys, this approach is not feasible for a college-wide assessment. In a "snapshot" assessment, as conducted, it is not possible to tease out the attitudinal differences that emanate from taking additional classes and student mindsets that existed beforehand and guided students to specific courses and programs. The weights of these two possibilities can be better estimated at the department level with additional surveys. However, comparisons of the attitudinal differences between students who have taken program-level STEM courses in specific disciplines to students who have not taken program-level classes show significant differences in all of the attitudinal categories (Figure 7). Therefore, emphasis should be placed on providing a more diverse and inclusive presentation of the history of scientific advancements, discoveries, and notable scientists.

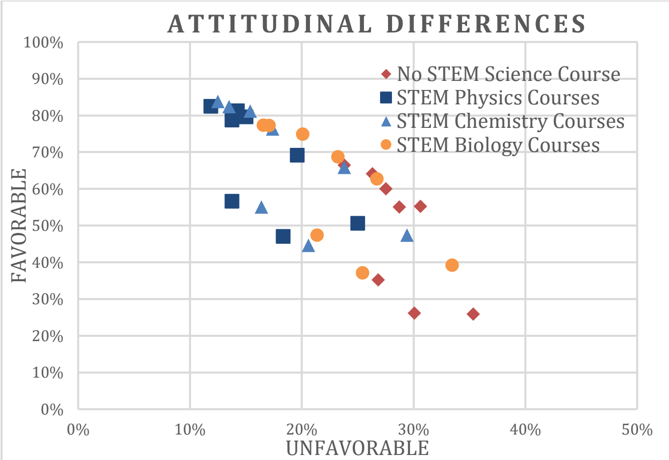

Figure 7 Attitudinal differences among the eight affective categories measured in from the CLASS questions for students taking program-level STEM course in three disciplines compared to students not taking any program level STEM cources.

In addition to examining attitudinal differences as they relate to academic history, this was also investigated in terms of ethnicity and gender. To be unambiguous, these results are not showing a causal relation between demographics and attitude about the sciences. They are, however, showing a gap in how typically under-represented students in the sciences view the sciences (Figures 8 and 9). While these exist in other institutions, they can have a larger impact at Harold Washington College due to the diverse student population.

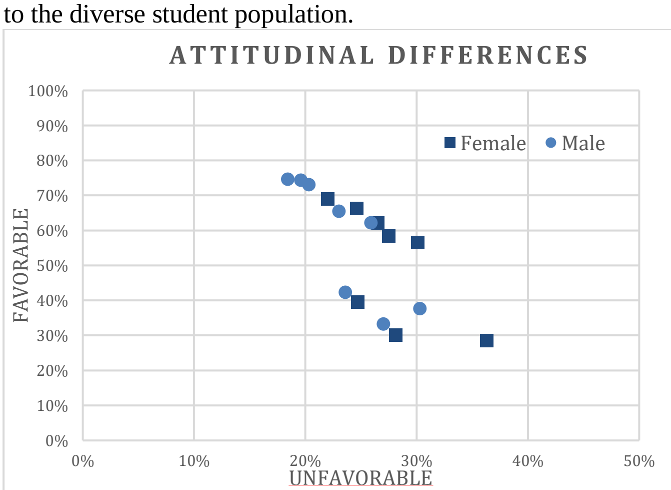

Figure 8 Attitudinal differences among the eight affective categories measured in the CLASS questions based on gender.

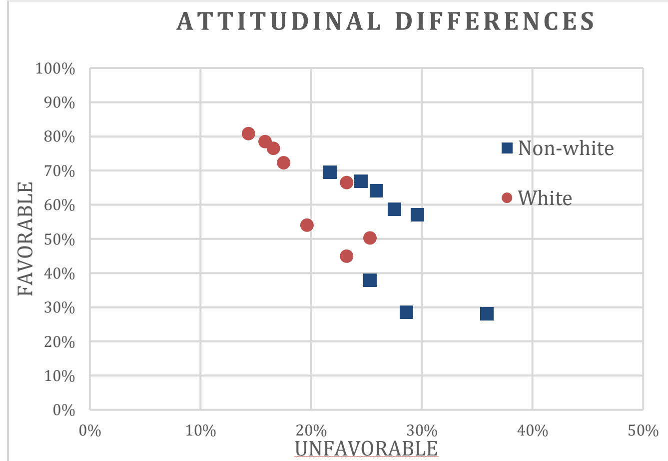

Figure 9 Attitudinal differences among the eight affective categories measured the CLASS questions based on ethnicity.

V. Discussion

The implementation of Openbook has provided this assessment with the most accurate, detailed, and exhaustive view of a students' academic histories. However, even with this data, it is still difficult to arrive at strong quantitative conclusions. The students entering Harold Washington College fall along a long spectrum of academic preparedness, access to recourses, and outside support. Their learning in the natural sciences is directly affected by the diversity of the curriculum, as well as their quantitative literacy, and reading comprehension. They enter our college at multiple points, and their academic paths are nonlinear. Due to the inherently nonlinear, multi-variant, highly correlated structure, it is imperative that caution be used when making generalizations regarding learning, and even more so when applying policy. While the data ascertained here elucidates our understanding of student learning in the Natural Sciences, its interpretation was only possible with multiple conversations with natural science faculty members. Without this context, this assessment would not have been possible.

While it was not possible to control for all the variables obtained with Openbook, it is clear our students are learning. The students completing STEM majors consistently perform statistically higher on this tool then the general body. When controlling for what appear to be the major factors, students who take general education natural science courses perform statistically better than those students who do not. The pedagogical, curricula, and administrative recommendations in this report are built on a strong foundation.

VI. Recommendations

In light of the results and experiences with this assessment, the Assessment Committee recommends the following:

- *II.A. General Education Goal and Departmental Structure:* Departments should consider designing pamphlets to better explain the courses within their department as well as their intending audiences to help students navigate their academic fields.
- *II.A. General Education Goal and Departmental Structure:* Departments should consider periodic reviews for their courses' student learning outcomes and confirming they align with their discipline's general education objectives.
- *II.B. Tool Development:* Conversation on how to shift outcomes from content-driven to process-driven should be explored.
- *II.C. Data Acquisition and Processing*: Electronic data acquisition, specifically Google Forms, is both efficient and effective, and its use should be continued.
- *II.C. Data Acquisition and Processing:* Students do not have reservations providing valid student identification, and, this should be used in extensive demographic questions to reduce survey lengths and testing fatigue.
- *IV.A. Performance Distributions Department Level*: Adoption of this tool for introductory topics courses in the natural sciences should be considered.
- *IV.A. Performance Distributions Department Level*: Unit-level assessment liaisons should continue to work with faculty members to ensure these objectives and outcomes are incorporated into the curriculum.
- *IV.B. Performance Distributions Academic Discipline Level*: Students may be entering subsequent classes with knowledge on how to complete a subset of scientific problems without understanding the principles that underlie them. It may be beneficial to assess and review these concepts.
- *IV.C. Correlation between Course History and Assessment Performance:* Assessing student learning solely by number of courses taken is not valid without addressing entrance point and course sequence.
- *IV.C. Correlation between Course History and Assessment Performance:* Students entering the college below credit level in mathematics and English are performing poorer on assessment outside of these disciplines. Support strategies should be considered to reduce this.
- *IV.C. Correlation between Course History and Assessment Performance:* Efforts to include more college-wide general education student learning outcomes in college courses should be considered.
- *IV.C. Correlation between Course History and Assessment Performance:* Allocating time to discuss similarities and differences between faculty member's academic disciplines could help provide a better understanding of the objectives of their disciplines.
- *IV.D. Attitudinal Shifts:* Emphasis should be placed on providing a more diverse and inclusive presentation of the history of scientific advancements, discoveries, and notable scientists.

VII. Acknowledgements

The Harold Washington College Assessment Committee sincerely thanks the students of the college for volunteering their time to complete the survey, faculty for volunteering their classes, and administration for supporting the committee with the necessary resources to make this whole project possible.

Appendix A

Biological Science Content Questions

- Which of the following is a correct sequence of structures that can be found in living things, proceeding from least to most complex?
- The basic unit of structure and function for all life is the
- Which of the following BEST describes the common flow of genetic information in a cell?
- Which of the following molecules store and transmit hereditary information?
- What is the relationship among DNA, genes and chromosomes?
- What is the relationship between mutations and cancer?
- A new family moved into your neighborhood. They have five children, all of whom are boys, and the mother is pregnant. Approximately, what is the probability that the new baby will be a boy?
- Dogs have 78 chromosomes in most cells of their body. How many chromosomes would you expect to find in a dog sperm or egg cell?
- When a plant-eating animal (herbivore) consumes a plant,
- When sunlight reaches a plant,

Appendix B

Physical Science Content Questions

- According to modern ideas and observations, what can be said about the location of the center of the Universe?
- As seen from your current location, when will an upright flagpole cast no shadow because the Sun is directly above the flagpole?
- What is the best definition for the scientific term "theory"?
- Below the outermost rocky shell of the Earth, it becomes
- Select the situation in which rain is most likely to form.
- Why is the sky blue?
- Which of the following best describes what happens when a pot of water boils?
- Above is the information from the periodic table for nitrogen. Based on this information, which statement is ALWAYS true?
- You put fresh batteries into a flashlight. Then you turn it on and leave it on until the bulb gradually dims and finally goes out. Which statement best describes the involvement of energy in this process?
- Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a

single story building at the same instant of time. The time it takes the balls to reach the ground below will be

Appendix C

Affective Questions

- A significant problem in learning science is being able to memorize all the information I need to know.
- I think about the science I experience in my life.
- Knowledge in science consists of many disconnected topics.
- I am not satisfied until I understand why something works the way it does.
- I do not expect equations to help my understanding of the ideas; they are just for doing calculations.
- If I get stuck on a science problem on my first try, I usually try to figure out a different way that works.
- Nearly everyone is capable of understanding science if they work at it.
- If I want to apply a method used for solving one scientific problem to another problem, the problems must involve very similar situations.
- I enjoy solving scientific problems.
- In science, mathematical formulas express meaningful relationships among measurable quantities.
- Learning science changes my ideas about how the world works.
- Reasoning skills used to understand scientific concepts can be helpful to me in my everyday life.
- Spending a lot of time understanding where formulas come from is unnecessary.
- To understand science, I sometimes think about my personal experiences and relate them to the topic being analyzed.
- When studying science, I relate the important information to what I already know rather than just memorizing it the way it is presented.