VIETNAM JOURNAL OF EDUCATION, 2021
Volume 5, issue 1, 52-63
https://doi.org/10.52296/vje.2021.45

ORIGINAL ARTICLE

Outcome Assessment of General Education 21st Century Competency Skills: Close the Loop Efforts in the United States

Nhung Thi Tuyet Pham

Hue University - University of Foreign Languages, Vietnam
Email: pttnhung.hufl@hueuni.edu.vn

Article History
Received: 5 January 2021
Accepted: 6 March 2021
Published: 30 March 2021

Keywords
Outcome assessment, general education, 21st century competency skills, higher education

ABSTRACT
Over the past several years, various stakeholders have continued to express concerns about the quality of U.S. higher education. Under the accountability and transparency pressures, institutions must provide evidence of student learning, especially the value of general education programs upon graduation. Therefore, a case study at a U.S. comprehensive university was conducted to assess five general education competency skills (written, oral, quantitative literacy, critical thinking and information literacy). To facilitate "close the loop" conversations with faculty and committees, in addition to descriptive analysis, the university disaggregated the assessment data in a non-traditional way by examining the relationship of student factors (race, year and college) and student learning. The researcher used ANCOVA and ANOVA to identify significant differences. Results indicated year and race were related to student outcomes, except for critical thinking skills. The researcher provided suggestions for use of the study's findings to close the loop in the general education program.

1. INTRODUCTION

Over the past several years, various individuals, organizations, and legislators have continued to express concerns about the quality of higher education in the United States. Those concerns have triggered legislation and requirements at the federal and state levels and by regional accreditors to assess and report on student learning (Jones, 2009; Nelson, 2014; Bassis, 2015). Therefore, U.S. colleges and universities have been asked to provide increasing evidence of transparency and accountability aimed at holding providers responsible for their performance through the disclosure of comparative results of programs, educational processes and outcomes to better inform prospective students and other stakeholders (McCormick, 2010; Liu, 2011). In 2005, the Spelling Commission report suggested that institutions be required to provide evidence of student learning outcomes (SLOs) for accountability (U.S. Department of Education, 2006). This report served as a push for revisiting outcomes assessment to demonstrate student success during their undergraduate experience.

Since then, there have been numerous studies on assessment methods to demonstrate evidence of student performance on 21st century competencies skills. There are four major approaches to assess students' 21st century skills: (a) embedding the assessment in general education programs, (b) examining the psychometric component of the assessment measures, (c) designing the assessment approach to provide learning gains or the "value-added" of students' competency skills, and (d) using multiple assessment measures to provide triangulated evidence of student 21st century competency outcomes to internal and external stakeholders. Most of the four assessment approaches use data from standardized exam to provide evidence of students' 21st century competency skills.

Beauchman and Waldenberger (2017) shared their experience with assessing a five-year assessment plan by using a course-embedded approach and then comparing student performance across the majors to facilitate a discussion with committees and departments. Al-Lail and Oudghiri (2016) not only used a rubric to assess institutional learning outcomes (ILOs) but also triangulated with indirect measures such as a market study to learn about the institution's reputation, a student survey and an employer survey. Martins et al. (2019) shared the assessment results from two measures—the Valid Assessment of Learning in Undergraduate Education (VALUE) rubric and the National Survey

of Student Engagement (NSSE). This research also emphasized the use of NSSE results to make interventions for better student engagement. Noticeably, all of these studies used the VALUE rubric and NSSE to provide descriptive evidence of student learning for accountability purposes. The VALUE rubric, as proposed by the American Association of Colleges and Universities (AAC&U), has a set of 16 learning outcomes (AAC&U 2015) to assess students' knowledge, skills and attitudes. VALUE rubrics serve as a supportive measure of authentic assessment to assess students' application of knowledge to address real problems (Boyles, 2012). In addition to descriptive statistics of student learning, Eisnaugle (2018) and Rear (2019) also looked further for additional factors that would have a relationship to student learning. Eisnaugle's (2018) studies indicated a significant correlation between course assignment and course GPA. Rear (2019) examined the relationship of GPA, admission and major on student learning.

The literature review showed that these previous studies included descriptive results but only limited empirical analysis of individual factors such as race, year, and college from authentic assessment data to examine how student factors relate to GE competency performance. To address the gap in the field, the purpose of this quantitative study was to disaggregate authentic assessment data in a non-traditional way to examine whether there were significant differences of race, year, college and their various interactions with five 21st century competencies while considering GPA as a covariate. The researcher hoped the assessment findings from this study could facilitate close the loop conversations with faculty and thereby improve teaching and learning as well as provide evidence of accountability.

2. LITERATURE REVIEW

In response to pressures from stakeholders to be accountable and transparent, institutions need to assess General Education (GE) programs and demonstrate the programs' value in contributing to student learning during undergraduate training. Research on GE assessment has included four approaches. The first research approach focuses on efforts to improve GE program by aligning course learning outcomes with GE outcomes (Galle & Galle, 2010; Scott & Fuess, 2011; Al-Lail & Oudghiri, 2016). This approach includes strategies to improve the effectiveness of GE assessment processes such as faculty engagement in assessment activities (Macdonald et al., 2014; Swarat & Wrynn, 2017; Faleski & Hand-Miller, 2017); leadership support of the GE assessment process (Al-Lail & Oudghiri, 2016; Swarat & Wrynn, 2017; Fulcher, Ames & Hawk, 2017); consistency, fairness and efficiency (Rust, Price & O'Donovan, 2003); transparency (Evans, 2013); and ways to communicate GE assessment results to internal and external committees (Macdonald et al., 2014; Faleski & Handley-Miller, 2017; Caspersen, Smeby & Olaf Aamodt, 2017). Research has been aimed at supporting faculty in GE assessment, which included principles for providing constructive feedback to students, such as e-assessment, peer assessment, and self-assessment (Caspersen et al., 2017); tips for avoiding bias in assessment of student work (Steinke & Fitch, 2017); and pedagogical strategies (e.g., a reflective piece to improve competencies learning) (Boyles, 2012; Atkinson & Lim, 2013).

The second research approach examines the psychometric component of assessment measures such as inter- and intra-rater reliability, rating consistency in authentic assessment (Hathcoat & Penn, 2012), construct validity of the rubric (Finley, 2011; Szafran, 2017; Gray, Brown & Connolly, 2017) and reliability of the national student survey (Wiewiora & Kowalkiewicz, 2019). Studies by Reddy and Andrade (2010), Banta and Palomba (2014) and Hack (2015) supported the concept that locally modified or created rubrics can produce valid and reliable results when colleges develop them responsibly.

The third research approach aims at accountability by seeking evidence of learning gain or value-added upon graduation (Liu, 2017). Eisnaugle (2018) assessed evidence to demonstrate students' value-added competencies in the social work discipline. Fulcher et al. (2017) assessed ethical reasoning by using a locally-built eight-dimension rubric, comparing data from freshmen and seniors. Fulcher et al. (2017) also used a local rubric to assess student competency, and Roohr et al. (2019) used a standardized exam to demonstrate the competency value-added. Roohr et al. (2019) also examined the impact of additional students' and institutional factors (GPA, admission and major) on students' critical thinking skills. Results indicated that institutions had an impact on the variation between freshmen and senior learning performance. Similarly, Demeter, Robinson, and Frederick (2019) collected data from a random sample of freshmen and seniors who took the Educational Testing Service (ETS) HEIghten exam and results showed student improvement in writing but not in critical thinking. Although the standardized exam is commonly used to provide evidence of accountability, Rear (2019) argued the standardized exam in critical thinking had low reliability and validity and failed to evaluate practical and real-life skills. He recommended using authentic and course-based assessment integral to the curriculum of a specific discipline. These studies all used the same method to collect data from freshmen and seniors, then compare for statistical difference, but the assessment measure was different. A standardized exam was commonly used to demonstrate evidence of value added as it is easy to make

a public comparison for accountability purposes (Edwel, 2009). However, the use of a rubric to assess a GE program is better fitted to the local context for quality improvement (Boyles, 2012; Carless, 2015).

The fourth research approach uses multiple assessment measures to provide evidence of student learning for stakeholders (Liu, 2017; Bruce, 2018). From an assessment perspective, the two major types of assessment measures are direct and indirect. Direct measures include the direct quantitative measurement of student work that requires students to display the knowledge, skills, and abilities they have learned. Indirect measures normally require students to reflect on how well they perform and mostly use surveys (Allen, 2004; Suskie, 2009). Institutions tend to use multiple assessment measures to triangulate for accurate results (Batini et al., 2009; Al-Lail & Oudghiri, 2016; Beauchman & Waldenberger, 2017; Liu, 2017; Martins et al., 2019). All the aforementioned studies provided descriptive measures of student performance.

3. RESEARCH METHODS AND RESULTS

3.1. Context

This research was based on assessment of a GE program at a regional comprehensive university in the Midwestern U.S. with an enrollment of about 12,000 undergraduate and graduate students. The GE program has always had the mission of providing students with foundational knowledge and skills that encompass all baccalaureate programs. In 2014, the university revised the GE program and required that courses be aligned with specific outcomes for the university's GE program and provide evidence that students have achieved the outcomes for the course and the program (Beauchman & Waldenberger, 2017; Bruce, 2018; Galle & Galle, 2010).

To assess the GE program, the university used multiple assessment measures: the general education assessment (GEA) administered by ETS, the NSSE and the Modified VALUE rubric. In the first year of reporting, all the data were aggregated in average by each competency and met the university benchmark. Therefore, the university did not take any specific actions to make improvements. Based on the findings of previous research studies' efforts to learn more about student learning in GE competencies, the university decided to examine the relationships of student factors (race, year, college) with the five competencies (written, oral, quantitative literacy, critical thinking and information literacy). To facilitate conversations on closing the loop with faculty and committees, the university decided to analyze data from the modified VALUE rubric (Rear, 2019). This decision demonstrated the institution's commitment to implementing course-based, authentic assessment that provides students with practical and real-life learning opportunities (Hathcoat & Penn, 2012; Boyles, 2012). Learner diversity (e.g., race) was factored in to achieve equity in assessment (Montenegro & Jankowski, 2017). Most importantly, course-embedded and authentic GE assessment provided the opportunity for faculty to have conversations about assessment results and strategies to close the loop (Boyles, 2012; Liu, 2017).

3.2. Study purpose and research questions

Individual factors to disaggregate the assessment of 21st century competencies in a standardized exam served as the framework to disaggregate authentic assessment for this study. The purpose of this quantitative study was to examine whether there were significant differences of race, year, college and their various interactions with five 21st century competencies while considering GPA as a covariate. The research addressed five research questions:

1. After controlling for GPA, are there differences in student performance in written communication by race, year and college?

2. After controlling for GPA, are there differences in student performance in oral communication by race, year and college?

3. After controlling for GPA, are there differences in student performance in quantitative literacy by race, year and college?

4. After controlling for GPA, are there differences in student performance in critical thinking by race, year and college?

5. After controlling for GPA, are there differences in student performance in information literacy by race, year and college?

3.3. Research methods

Participants were students from a Midwest Comprehensive University in the U.S. which had completed GE courses in written, oral, quantitative literacy, critical thinking and information literacy as part of a GE program. The total number of participants for all five competencies was 3261. See Table 1 for participant demographics.

Table 1. Participant Number and Demographics

	Written Communication	Oral Communication	Quantitative Literacy	Critical Thinking	Information Literacy
			Race		
Black	*101*	*49*	*128*	*53*	*78*
White	*520*	*361*	*748*	*353*	*649*
Other	*62*	*26*	*63*	*29*	*41*
			Year		
Freshman	130	68	136	72	87
Sophomore	240	99	252	128	207
Junior	163	126	289	105	250
Senior	150	143	262	130	224
			College		
Social Sciences	93	90	112	73	129
Education	74	73	164	48	90
Science & Technology	347	116	446	235	315
Business	169	157	217	79	234
Total	683	436	939	435	768
Participation rate	48%	42%	76%	33%	63%

3.3.1. Instrument

The university used a course-embedded assessment approach to collect data for GE assessment. Modified VALUE rubrics were the assessment tools. When modifying the five rubrics of oral, written, quantitative, critical thinking and information literacy, the university just simplified the rubrics by choosing fewer dimensions and mostly kept the original language in the rubric. Therefore, the rubrics still demonstrated validity (Reddy & Andrade, 2010; Banta & Palomba, 2014; Hack, 2015). The five rubrics have three to four dimensions with scores from N/A, one (novice) to four (highly developed).

3.3.2. Materials

Data were collected in Fall 2017 and Fall 2018. Any courses aligned with a skill-based competency were required to submit an assignment that assessed all the components in the modified VALUE rubric. The university sent a link from an assessment management system (AMS) to a learning management system (LMS) to collect artifacts from more than 150 courses. The university encouraged faculty to submit a current and graded assignment into the system to reduce additional workload and have a high participation rate. The artifacts were rated by the instructor teaching the class. In order to obtain the data for student performance in five competencies by race, year and college, the university generated individual competency reports in the AMS by filtering certain variables. The year variable was chosen to provide value-added results (Hawk, 2017; Eisnaugle, 2018; Demeter et al., 2019; Roohr et al., 2019). Aggregated data by race could provide faculty additional information about equity in assessment (Montenegro & Jankowski, 2017). Also, to facilitate the communication of assessment results to different committees and close the loop, the university aggregated data by college. The assessment staff thought the analysis of additional student factors could provide more specific data to internal committees about the evidence of student learning and possible actions for improvement. This data also served as evidence for external accountability such as regional accreditation.

3.3.3. Design and Statistical Procedure

The three categorical independent variables (IVs) were year with four levels (Freshman, Sophomore, Junior, and Senior), race with three levels (Black, White, and Other), and college with four levels (College of Arts, Humanities and Social Sciences [CAHSS], College of Education [COE], College of Health Science and Technology [CHST], and College of Business and Professional Studies [CBPS]). The covariate variable was GPA. The dependent variables (DVs) were student performance in written communication, oral communication, quantitative literacy, critical thinking, and information literacy. A series of analyses were conducted to compare student performance by race, year, and college while controlling GPA as a covariate. A three-way analysis of covariance (ANCOVA) using the Statistical Package for Social Sciences (SPSS) and a three-way analysis of variance (ANOVA) using SPSS were also conducted to evaluate the interaction between year, race and college on GE student performance without controlling for GPA.

This research had some limitations. The study only collected one data point for five competencies; therefore, it might not provide the most accurate learning gains of student performance. Although the research used individual factors from assessment of standardized exam to analyze authentic assessment, there was no strong theoretical framework to choose college as the IV for the study. As the research design was bounded by a case study, the assessment results should be carefully interpreted in that context. Further research should be conducted to collect multiple data points and increase the number of raters to improve the reliability.

3.4. Research results

3.4.1. Statistical assumptions

To ensure the data had been appropriately collected for ANCOVA and ANOVA statistical analyses, the researcher checked 10 statistical assumptions. The research met the assumptions of one continuous DV (student performance), three categorical IVs (year, race, college), and one continuous covariate (GPA). The sample size for each competency was more than 30. Based on sample size, the researcher applied Central Limit Theorem (Field, 2009) to meet the assumptions of homogeneity of regression slopes, homoscedasticity, homogeneity of variances, and normality. There were no outliers in the data greater than ±3 standard deviations.

3.4.2. Written communication

Three-way ANCOVA

There was not a statistically significant interaction between race, year, and college on Written Communication, whilst controlling for GPA $F(13, 639)$, $p = .386$, partial $\eta 2 = .021$. Therefore, an analysis of the main effects for race, year, and college was performed. Seniors had higher scores than freshmen in written communication while controlling for GPA $F(3,639)$, $p = .002$, $\eta 2 = .023$. Means, adjusted means, standard deviations and standard errors are presented in Table 2.

Table 2. Written Communication

College of Art Humanity												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.0	2.0	2.1	1.5	1.7	1.2	1.5	1.9	1.0	2.0	2.1	3.3
(SD)	(1.4)	(.94)	(.70)	(1.2)	(.73)	(.60)	(.71)	(1.1)	(.00)	(.71)	(1.0)	--
Madj	2.1	2.2	2.5	1.6	1.7	1.2	1.6	1.8	1.1	--	2.0	3.3
(SE)	(.42)	(.24)	(.49)	(.41)	(.15)	(.41)	(.59)	(.18)	(.59)	--	(.28)	(.83)

College of Education												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.0	1.7	2.1	1.3	1.7	1.3	2.1	2.5	2.0	--	2.7	--
(SD)	--	(1.0)	--	(.66)	(.89)	(.47)	--	(1.1)	(1.4)	--	(.95)	--
Madj	1.1	1.9	--	1.3	1.7	1.3	--	2.4	1.9	--	2.6	--
(SE)	(.83)	(.24)	--	(.42)	(.16)	(.58)	--	(.19)	(.58)	--	(.32)	

College of Science & Technology												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.5	1.7	2.0	1.5	1.5	2.1	2.1	2.5	1.5	2.0	2.7	2.1
(SD)	(.56)	(.95)	--	(.77)	(.73)	(.80)	(.65)	(.93)	(.60)	(.71)	(.89)	(1.2)
Madj	1.5	1.8	1.9	1.6	1.5	2.1	2.1	2.4	1.5	2.0	2.6	2.0
(SE)	(.20)	(.13)	(.83)	(.17)	(.10)	(.34)	(.28)	(.12)	(.28)	(.26)	(.10)	(.22)

College of Business												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.7	1.6	2.1	1.4	1.8	1.8	1.6	2.5	1.8	2.1	2.0	--
(SD)	(.82)	(.72)	(.60)	(.78)	(.63)	(1.0)	(.68)	(.70)	(.20)	(1.0)	(.83)	

Madj	1.9	1.8	2.3	1.4	1.8	1.8	1.6	1.6	1.7	2.1	2.0	1.8
(SE)	(.42)	(.19)	(.38)	(.26)	(.13)	(.42)	(.31)	(.13)	(.50)	(.50)	(.17)	(.37)

Three-way ANOVA

There was not a statistically significant three-way interaction between year, race and school, $F(13, 640) = 1.030$, $p = .420$. There was a statistically significant difference by year on Written Communication, while not controlling for GPA $F(3, 640) = 5.16$, $p = .002$. There was no statistically significant difference by race, $F(2, 640) = .828$, $p = .438$, or with school $F(3, 640) = .563$, $p = .640$ on Written Communication. There was also no statistically significant interaction between race and year $F(6, 640) = .583$, $p = .744$, between college and year $F(9, 640) = 1.13$, $p = .337$, and between race and college $F(6, 640) = .446$, $p = 848$ on Written Communication.

3.5. Oral Communication

Three-way ANCOVA

There was no statistically significant interaction between race, year, and school on Oral Communication, whilst controlling for GPA $F(11, 394)$, $p = .726$, partial $\eta 2 = .020$. Therefore, an analysis of the main effects for race, year, and college was performed. Seniors had higher scores than freshmen in oral Communication, whilst controlling for GPA, $F(3, 394)$, $p = .000$, partial $\eta 2 = .044$. Means, adjusted means, standard deviations and standard errors are presented in Table 3.

Table 3. Oral Communication

College of Art & Humanity												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.2	2.4	3.0	2.1	2.6	--	1.6	2.8	--	3.0	3.3	3.5
(SD)	(.88)	(1.3)	--	(1.1)	(1.0)	--	(.95)	(1.1)	--	--	(.64)	--
Madj	2.3	2.6	3.2	2.1	2.6	--	1.8	2.7	--	3.1	3.3	3.5
(SE)	(.51)	(.24)	(.88)	(.44)	(.21)	--	(.44)	(.17)	--	(.88)	(.21)	(.90)

College of Education												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.0	20	3.0	2.9	2.7	--	2.8	3.0	1.5	--	3.5	--
(SD)	(.50)	(1.4)	--	(.59)	(.80)	--	(.71)	(.90)	--	--	(.70)	--
Madj	2.4	2.2	--	3.0	2.6	--	2.7	2.9	1.5	--	3.3	--
(SE)	(.52)	(.29)	--	(.62)	(1.7)	--	(.62)	(.20)	(.88)	--	(.33)	--

College of Science & Technology												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.6	2.0	--	1.8	2.5	1.9	1.9	2.6	2.0	2.9	3.2	3.4
(SD)	(.53)	(1.1)	--	(.24)	(.98)	(.77)	(.55)	(.76)	--	(.80)	(.87)	(.93)
Madj	1.7	2.1	--	1.9	2.4	2.0	2.0	2.5	1.8	2.9	3.2	3.3
(SE)	(.62)	(.30)	--	(.43)	(.22)	(.50)	(.44)	(.16)	(.90)	(.40)	(.15)	(.40)

College of Business												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.3	1.2	2.2	2.0	1.9	2.9	2.0	2.6	2.8	3.1	3.3	2.4
(SD)	(.63)	(.84)	(.70)	(.50)	(.71)	(1.2)	(.55)	(.91)	(.40)	(.90)	(.70)	(1.5)
Madj	1.5	1.5	2.2	2.0	1.9	2.9	2.0	2.5	2.8	3.1	3.2	2.4
(SE)	(.44)	(.24)	(.40)	(.50)	(.20)	(.62)	(.90)	(.20)	(.62)	(.33)	(.11)	(.40)

Three-way ANOVA

There was no statistically significant three-way interaction between year, race and college, $F(11, 395) = .725$, $p = .715$. However, there was a statistically significant difference by year on Oral Communication, $F(3, 395) = 7.800$,

$p < .001$. There was no statistical significance with race $F(2, 395) = 1.846$, $p = .159$, and with college, $F(3, 395) = .821$, $p = .483$ on Oral Communication. There was also no statistically significant interaction between year and race, $F(6, 395) = .619$, $p = .715$, between year and college $F(9, 395) = .587$, $p = .808$, and between race and college, $F(6, 395) = .890$, $p = .502$ on Oral Communication.

3.6. Quantitative Literacy

Three-way ANCOVA

There was no statistically significant interaction between race, year, and school on Quantitative Literacy, whilst controlling for GPA, $F(16, 875)$, $p = .472$, partial $\eta 2 = .018$. Therefore, an analysis of the main effects for race, year, and college was performed. Seniors had higher scores than freshmen and Black students scored lowest in quantitative Literacy, $F(6, 875)$, p = .015, partial $\eta 2 = .018$. Means, adjusted means, standard deviations and standard errors are presented in Table 4.

Table 4. Quantitative Literacy

College of Art & Humanity												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.3	2.1	2.3	2.5	2.7	--	1.6	2.6	--	1.9	2.4	2.2
(SD)	--	(.74)	--	(.71)	(1.0)	--	(.88)	(.86)	--	(.80)	(.91)	(1.0)
Madj	2.4	2.3	2.3	2.5	2.7	--	1.7	2.6	--	1.9	2.3	2.1
(SE)	(.80)	(.23)	(.80)	(.57)	(.20)	--	(.40)	(.13)	--	(.33)	(.16)	(.33)

College of Education												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.0	2.1	4.0	2.7	2.6	2.2	2.1	2.5	3.0	1.5	2.7	4.0
(SD)	(.00)	(.67)	--	--	(.66)	(.71)	(.32)	(.65)	--	(.71)	(.73)	--
Madj	2.3	2.2	4.0	2.5	2.5	2.2	2.1	2.4	2.9	1.5	2.6	3.9
(SE)	(.57)	(.20)	(.80)	(.56)	(.12)	(.56)	(.40)	(.11)	(.80)	(.56)	(.13)	(.80)

College of Science & Technology												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.9	2.1	1.8	2.3	2.4	1.6	2.3	2.4	2.8	2.0	2.5	2.6
(SD)	(1.1)	(.67)	(1.6)	(.74)	(.84)	(.70)	(.80)	(.75)	(1.5)	(.47)	(.82)	(.97)
Madj	2.0	2.3	2.0	2.3	2.3	1.6	2.3	2.3	2.8	2.0	2.4	2.6
(SE)	(.23)	(.13)	(.46)	(.21)	(.08)	(.27)	(.16)	(.09)	(.23)	(.24)	(.08)	(.21)

College of Business												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.0	2.3	2.4	2.1	2.3	1.0	1.6	2.3	2.2	2.6	2.1	3.0
(SD)	(.80)	(.87)	(1.3)	(.80)	(1.0)	(1.4)	(.64)	(.77)	(1.5)	(.76)	(.87)	(3.0)
Madj	2.2	2.5	2.9	2.2	2.2	.84	1.9	2.3	2.1	2.7	2.1	2.9
(SE)	(.29)	(.16)	(.47)	(.30)	(.11)	(.56)	(.19)	(.12)	(.36)	(.27)	(.12)	(.80)

Three-way ANOVA

There was no statistically significant three-way interaction between year, race and college, $F(16, 876) = .1.100$, $p = .350$; however, there was a statistically significant interaction between year and race $F(6, 876) = 2.843$, $p = .010$ on Quantitative Literacy. There was no statistically significant interaction between year and college, $F(9,876) = .583$, $p = .531$, and between race and college $F(6,876) = 1.37$, $p = .223$ on Quantitative Literacy. There was also no statistically significant difference by year $F(3, 876) = 1.02$, $p = .384$ and by college $F(3, 876) = 1.85$, $p = .14$ on Quantitative Literacy.

3.7. Critical Thinking

Three-way ANCOVA

There was no statistically significant interaction between race, year, and school on Critical Thinking, whilst controlling for GPA, $F(9, 395)$, $p = .871$, partial $\eta 2 = .011$. Therefore, an analysis of the main effects for race, year, and college was performed. There was no statistically significant interaction between year and race on Critical Thinking, $F(6, 395)$, $p = .480$, partial $\eta 2 = .014$. Means, adjusted means, standard deviations and standard errors are presented in Table 5.

Table 5. Critical Thinking

College of Art & Humanity												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	2.1	2.2	--	4.0	2.6	--	2.1	2.6	--	2.5	2.6	2.5
(SD)	(.13)	(.94)	--	--	(1.1)	--	(.18)	(1.1)	--	(1.7)	(1.1)	(2.1)
Madj	2.2	2.2	--	4.1	2.6	--	2.2	2.5	--	2.6	2.5	2.5
(SE)	(.53)	(.30)	--	(1.1)	(.24)	--	(.80)	(.32)	--	(.53)	(.27)	(.80)

College of Education												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	--	1.4	2.8	--	2.3	.80	1.6	1.9	--	3.3	2.3	--
(SD)	--	(.53)	--	--	(.99)	--	(.53)	(1.0)	--	--	(1.0)	--
Madj	2.2	1.5	2.7	--	2.2	.63	1.6	1.9	--	3.4	2.3	--
(SE)	(.54)	(.80)	(1.1)	--	(.34)	(1.1)	(.80)	(.27)	--	(1.1)	(.30)	--

College of Science & Technology												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.8	2.1	--	1.7	2.2	2.5	1.4	2.0	1.6	2.7	2.3	2.5
(SD)	(1.0)	(.81)	--	1.2	1.1	1.5	1.4	2.0	1.6	2.7	2.3	2.5
Madj	1.8	2.1	--	1.8	2.2	2.5	1.5	2.0	1.7	2.7	2.3	2.5
(SE)	(.40)	(.20)	--	(.404)	(.13)	(.50)	(.38)	(.16)	(1.1)	(.48)	(.15)	(.53)

College of Business												
Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.0	1.1	--	2.3	1.8	2.1	2.3	2.0	2.5	2.0	2.3	3.0
(SD)	--	(.40)	--	(.36)	(1.2)	(.63)	--	(1.3)	--	(1.2)	(1.0)	(1.3)
Madj	1.2	1.2	--	2.3	1.7	2.1	2.4	2.0	2.5	1.9	1.8	2.8
(SE)	(1.1)	(.30)	--	(.61)	(.34)	(.50)	(1.1)	(.30)	(1.1)	(.41)	(.24)	(.53)

Three-way ANOVA

There was no statistically significant three-way interaction between year, race and college on Critical Thinking, $F(11, 397) = .725$, $p = .714$. Also, there was no statistical significance with year $F(3, 397) = 1.73$, $p = .161$, with race, $F(2, 397) = .126$, $p = .161$, and with college on Critical Thinking, $F(3, 397) = .338$, $p = .798$. There was also no statistically significant interaction between year and race $F(6,397) = .985$, $p = .435$, between year and college $F(9, 397) = .796$, $p = .435$, and between race and college on Critical Thinking $F(6, 397) = 1.13$, $p = .343$.

3.8. Information Literacy

Three-way ANCOVA

There was no statistically significant interaction between race, year, and school on Information Literacy, whilst controlling for GPA, $F(14, 723)$, $p = .306$, partial $\eta 2 = .022$. Therefore, an analysis of the main effects for race, year, and college was performed. There was a statistically significant difference by race on Information Literacy $F(2, 723)$, $p = .024$, partial $\eta 2 = 0.10$. Means, adjusted means, standard deviations and standard errors are presented in Table 6.

Table 6. Information Literacy

College Art Humanity

Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	--	1.9	1.3	2.2	2.1	2.3	2.8	2.2	2.0	1.6	2.2	--
(SD)	--	(.89)	(.00)	(1.0)	(.86)	(2.3)	(1.6)	(.64)	(.14)	(.58)	(.82)	--
Madj	--	2.3	1.8	2.3	2.1	2.3	2.7	2.2	2.0	1.6	2.2	--
(SE)	--	(.30)	(.55)	(.31)	(.15)	(.53)	(.54)	(.12)	(.54)	(.44)	(.12)	--

College of Education

Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	.70	1.8	2.0	2.0	1.9	--	2.0	2.4	4.0	--	2.4	2.0
(SD)	--	1.1	--	--	(.38)	--	--	(.88)	--	--	(.88)	--
Madj	.97	2.2	1.8	2.2	1.8	--	1.9	2.3	3.8	--	2.34	2.1
(SE)	(.80)	(.28)	(.76)	(.76)	(.16)	--	(.76)	(.13)	(.76)	--	(.20)	(.76)

College of Science Technology

Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	White	Other	Black	White	Other	Black	White	Other
M	1.7	2.0	2.3	1.7	2.2	1.8	2.2	2.2	2.1	2.1	2.3	1.9
(SD)	(.67)	(.81)	--	(.50)	(.75)	(.33)	(1.1)	(.72)	(.88)	(.90)	(.80)	(.35)
Madj	1.8	2.2	2.4	1.7	2.3	2.0	2.2	2.1	2.0	2.1	2.2	1.9
(SE)	(.34)	(.17)	(.80)	(.21)	(.10)	(.30)	(.29)	(.09)	(.38)	(.23)	(.08)	(.27)

College of Business

Year	Freshman			Sophomore			Junior			Senior		
Race	Black	White	Other	Black	Race	Black	White	Other	Black	Race	Black	White
M	1.5	1.9	1.7	1.5	M	1.5	1.9	1.7	1.5	M	1.5	1.9
(SD)	(.84)	(.81)	(.47)	(.79)	(SD)	(.84)	(.81)	(.47)	(.79)	(SD)	(.84)	(.81)
Madj	1.9	2.3	2.0	1.6	Madj	1.9	2.3	2.0	1.6	Madj	1.9	2.3
(SE)	(.32)	(.16)	(.39)	(.29)	(SE)	(.32)	(.16)	(.39)	(.29)	(SE)	(.32)	(.16)

Three-way ANOVA

There was no statistically significant three-way interaction between year, race and college, $F(14, 723) = 1.09$, $p = .367$ on Information Literacy. There was a statistically significant difference by year, $F(3, 724) = 3.28$, $p = .021$, and with race, $F(2, 724) = 4.9$, $p = .008$. There was no statistically significant difference with college on Information Literacy, $F(3, 724) = .022$, $p = .996$. There was no statistically significant interaction between year and race, $F(6, 724) = 1.07$, $p = .379$, between year and college $F(9, 724) = .998$, $p = .440$, and between race and college on Information Literacy, $F(6, 724)$, $p = 1.92$.

4. DISCUSION AND CONCLUSION

The research results indicated that seniors scored higher than freshmen in written, oral communication competencies and quantitative literacy when controlling and not controlling GPA. In addition, white students performed better in quantitative literacy than black students when controlling and not controlling GPA. There were no significant differences in student performance in critical thinking over the years when controlling and not controlling GPA. There were significant differences in student performance in information literacy by race when controlling GPA, and there were significant differences in student performance in information literacy by year when not controlling GPA. This meant GPA did not relate to student learning in all the competencies except for critical thinking and information literacy. There was no difference in student performance in all competencies by college. This finding could be explained by the fact that many first-year students had not identified their major or even chosen a major.

The findings of this research study demonstrated value-added in student learning from freshmen to seniors in four competencies. Value-added is a major assessment measure to demonstrate learning gains across the years in a

university (Fulcher, Ames & Hawk 2017; Rear 2019 and Demeter, Robinson & Frederick 2019). Of the research studies about value-added, only Fulcher et al. (2017) used a local rubric to examine the learning gain of ethical reasoning data collected from freshmen and seniors. Although the current study collected data from the first year, the sample pool also included all years. The lead GE faculty in each competency explained in the GE assessment discussion meeting that some students did not take GE courses until their senior year. This finding indicated the GE committee needed to revisit the current GE program, especially the GE course sequence and course requirement in the recertification process since about 30% of the students did not take GE courses until their senior year.

Although this research used authentic assessment to assess critical thinking, the findings of no significant difference in the standardized exam of critical thinking was the same as results from the study by Demeter et al. (2019). This result provided some insights for the internal committees to make improvements in the GE critical thinking skill courses. For example, revisit the assignment design of critical thinking skills to ensure it allows students to demonstrate their knowledge and skills at a highly developed level in the rubric since about 20% to 30% of the student population were seniors. Assignment design is also a good approach to engaging faculty to improve this skill (Bailey & McDevitt, 2019) since the descriptive analysis demonstrated that about 30% of the artifacts in critical thinking skills scored N/A in the dimension of "argumentation" in the rubric. Noticeably, GPA did not relate to student learning in all competencies. This result supports the necessity of assessment since students' grades do not reflect accurately the students' knowledge and skills upon their graduation (Suskie, 2009). Therefore, higher education institutions need to have an assessment system that provides evidence of student learning, especially student learning gained in addition to traditional grades in the transcript, to demonstrate accountability to stakeholders.

The literature has mostly provided descriptive assessment results for the purposes of accountability (Batini et al., 2009; Al-Lail & Oudghiri, 2016; Beauchman & Waldenberger, 2017; Liu, 2017; Martins et al., 2019). Only limited research has been published that used data disaggregation on the difference of institutional and individual factors on student competency performance in authentic assessment and efforts to use the authentic assessment results in "close the loop" conversations with faculty and internal committees for quality improvement. Student factors such as race and year address equity in assessment (DeSantis, 2020). Faculty teaching quantitative literacy and information literacy should be mindful of diverse learners in the equity of their instruction and assessment (Montenegro & Jankowski, 2017). Results of this study indicated that some Black students and others might need more support than white students in quantitative literacy and information literacy classes. Most importantly, the detailed results from this formative assessment provided the university with opportunities to make immediate interventions to help students perform better at the senior level. Also, to ensure the seniors have reinforced learning opportunities for these skills in the discipline, the university should assess student performance at the senior level such as capstone courses or courses in the discipline to provide stronger evidence of learning gain (Drisko, 2014)

This research has two implications for other higher education institutions. First, there is a need to address equity in assessment of student outcomes, and the disaggregation in this research also responds to the national call to provide better evidence of student learning. Second, the analysis of individual factors provides institutions, especially the faculty and universities' committees such as general education committee and university assessment committees, discussion opportunities to identify potential improvements to their GE program.

REFERENCES

Al-Lail, H. J., & Oudghiri, H. (2016). *Assessment of student learning outcomes based on institutional core values*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Allen, M. (2004). *Assessing academic programs in higher education*. Boltan, MA: Anker Publishing Company, Inc.

Atkinson, D., & Lim, S. (2013). Improving assessment processes in higher education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology*, *29*(5), 651-666. doi:10.14742/ajet.526

Bailey, S. &McDevitt, J. (2019). Palo Alto College: Faculty and Student Engagement through Assignment Design. *Case Studies*. Urbana, IL: National Institute for Learning Outcomes Assessment.

Banta, T. W., & Palomba, C. A. (2014). *Assessment essentials: planning, implementing, and improving assessment in higher education* (2nd ed.). San Francisco, CA: Jossey-Bass.

Bassis, M. (2015). A Primer on the transformation of higher education in America. Retrieved from http://www.learningoutcomeassessment.org/documents/BassisPrimer.pdf

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 6-20. doi:10.1145/1541880.1541883

Beauchman, M., & Waldenberger, S. (2017). Assessing general education: Identifying outcomes, data analysis, and improvements. *Assessment in Practice*. Urbana, IL: National Institute for Learning Outcomes Assessment.

Boyles, T. (2012). 21st century knowledge, skills, and abilities and entrepreneurial competencies: a model for undergraduate entrepreneurship education (Report). *Journal of Entrepreneurship Education*, *15*.

Bruce. R. T. (2018), Assessment in Action: Evidence-Based discussions about teaching, learning, and curriculum, *New Directions for Teaching and Learning*, *10*(2), 1-7.D. https://doi.org/10.1002/tl.20260

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963-976. doi:10.1007/s10734-014-9816-z

Caspersen, J., Smeby, J-C., & Olaf Aamodt, P. (2017). Measuring learning outcomes. *European Journal of Education*, *52*(1), 1-11. doi:10.1111/ejed.12205

Demeter, E., Robinson, C., & Frederick, J. G. (2019). Holistically assessing critical thinking and written communication learning outcomes with direct and indirect measures. *Research and Practice in Assessment*, *14*.

DeSantis, M. (2020). St. Edward's University: Rethinking general education to support social justice. Urbana, IL: University of Illinois and Indiana University, *National Institute for Learning Outcomes Assessment, Council for the Advancement of Standards in Higher Education, and Campus Labs*. Retrieved at https://www.learningoutcomesassessment.org/wpcontent/uploads/2020/09/EquityCase_St_Edwards.pdf

Eisnaugle, E. (2018). *Program evaluation of general education learning outcome assessment*. ProQuest LLC, Ed.D. Dissertation, Wingate University.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, *83*(1), 70-120. doi:10.3102/0034654312474350

Ewell, P. (2009). *Assessment, accountability, and improvement: Revisiting the tension*. (Occasional Paper No.1), Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from http://www.learningoutcomeassessment.org/documents/PeterEwell_005.pdf

Faleski, M., & Handley-Miller, B. (2017). *General education at Delta College: Guided conversations toward relevant assessment*. (Assessment in Practice), Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Field, A. (2009). *Discovering statistics using SPSS*. London, England: SAGE.

Finley, A. P. (2011). How reliable are the VALUE rubrics? *Peer Review*, *13*(4), 31-33.

Fulcher, K. H., Ames, A. J., & Hawk, W. J. (2017). *Ethical reasoning: Assessing a critical thinking skill*. (Assessment in Practice), Urbana, IL: National Institute for Learning Outcomes Assessment.

Galle, J. K., & Galle, J. (2010). Building an integrated student learning outcomes assessment for general education: Three case studies. *New Directions for Teaching and Learning*, *121*. doi:10.1002/tl.390

Gray, J., Brown, M., & Connolly, J. (2017). Examining construct validity of the quantitative literacy VALUE rubric in college-level STEM assignments. *Research & Practice in Assessment*, *12*.

Hack, C. (2015). Analytical rubrics in higher education: a repository of empirical data. *British Journal of Educational Technology*, *46*(5), 924-927. doi:10.1111/bjet.12304

Hathcoat, J., & Penn, J. (2012). Generalizability of Student Writing across Multiple Tasks: A Challenge for Authentic Assessment. *Research & Practice in Assessment*, *7*, 16-28.

Jones, D. A. (2009). Higher Education Assessment—Who Are We Assessing, and For What Purpose? Retrieved September 27, 2018, from https://www.aacu.org/publications-research/periodicals/higher-education-assessment%E2%80%94who-are-we-assessing-and-what-purpose

Liu, O. (2011). Outcomes Assessment in Higher Education: Challenges and Future Research in the Context of Voluntary System of Accountability, (Report). *Educational Measurement: Issues and Practice*, *30*(3), 2-9. doi:10.1111/j.1745-3992.2011.00206.x

Liu, O. (2017). Ten Years After the Spellings Commission: From Accountability to Internal Improvement, (Report). *Educational Measurement: Issues and Practice*, *36*(2), 34-41. doi:10.1111/emip.12139

Macdonald, S., Williams, L., Lazowski, R., Horst, S., & Barron, K. (2014). Faculty Attitudes toward General Education Assessment: A Qualitative Study about Their Motivation. *Research & Practice in Assessment*, *9*.

Martins, M. J., Reis Marques, T., Laíns, I., & Leal, I. (2019). The National Student Survey: validation in Portuguese medical students. *Assessment & Evaluation in Higher Education, 44*(1), 66-79. doi:10.1080/02602938.2018.14755470

McCormick, A. C. (2010), "Here's looking at you: Transparency, institutional self-presentation, and the public interest". *Change*, *42*(6), 35-43, doi:10.1080/00091383.2010.523406

Montenegro, E., & Jankowski, N. A. (2017). *Equity and assessment: Moving towards culturally responsive assessment*. (Occasional Paper No. 29). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Nelson, C. (2014). Assessing assessment. Retrieved from https://www.insidehighered.com/views/2014/11/24/essay-criticizes-state-assessment-movement-higher-education

Rear, D. (2019). One size fits all? The limitations of standardized assessment in critical thinking. *Assessment & Evaluation in Higher Education*, *44*(5), 664-675. doi:10.1080/02602938.2018.1526255

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*(4), 435-448. doi:10.1080/02602930902862859

Roohr, K., Olivera-Aguilar, M., Ling, G., Rikoon, S. (2019). A multi-level modeling approach to investigating students' critical thinking at higher education institutions. *Assessment & Evaluation in Higher Education*, *44*(2), 1-15. https://doi.org/10.1080/02602938.2018.

Rust, C., Price, M., & O'Donovan, B. (2003). Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes. *Assessment & Evaluation in Higher Education*, 28(2), 147-164. doi:10.1080/02602930301671

Steinke, P., & Fitch, P. (2017). Minimizing Bias When Assessing Student Work. *Research & Practice in Assessment, 12*.

Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.), San Francisco: Jossey-Bass.

Swaffield, S. (2011). Leading assessment for learning. *International handbook of leadership for learning*, 1047-1065, doi:10.1007/978-94-007-1350-5_57

Swarat, S. & Wrynn, A. M. (2017). *Assessment with Benefits: Faculty engagement and community building through GE assessment*. (Assessment in Practice), Urbana, IL: National Institute for Learning Outcomes Assessment.

Szafran, R. (2017). The miscalculation of interrater reliability: A case study involving the AAC&U VALUE rubrics. *Practical Assessment, Research & Evaluation*, *22*(11), 1-7.

The HEIghten® Outcomes Assessment Suite (2021). Retrieved from https://www.ets.org/heighten

U.S. Department of Education (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, DC: Author, doi:10.1086/512957

Wiewiora, A., & Kowalkiewicz, A. (2019). The role of authentic assessment in developing authentic leadership identity and competencies. *Assessment & Evaluation in Higher Education*, 44(3), 415-430. doi:10.1080/02602938.2018.1516730