International Journal of Learning, Teaching and Educational Research
Vol. 11, No. 1, pp. 212-226, April 2015

The Value-Added Assessment of Higher Education learning: The case of Nagoya University of Commerce and Business in Japan

Hiroshi Ito Surname and Nobuo Kawazoe
Nagoya University of Commerce and Business
Nisshin, Japan

Abstract. Assessment of higher education learning has been considered increasingly important. One of the current trends in this field is the value-added assessment—how much students learn during a certain period of time at university. In the United States, for example, Arum and Roksa (2011) conducted a large-scale assessment on second-year university students' learning with the Collegiate Learning Assessment (CLA) to examine how much university students improved generic skills during the first two years of higher education. Findings suggested that they did not improve much. The researchers concluded that the poor result was attributed to the fact that American university students on average study only 12 hours a week. In Japan, the situation may be even worse as Japanese university students on average study 3.5 hours, much less than their cohorts in the United States. However, studies on the value-added assessment of Japanese university students' learning are scarce. With the Progress Report on Generic Skills (PROG), an assessment tool similar to CLA, as well as interviews with students who took PROG, this study quantitatively examines how much students improved generic skills at a Japanese university during the first two years of higher education. The findings show that as was the case of the US peers, Japanese university students in this study did not improve their generic skills very much in the first two years of higher education. This study also qualitatively explores possible reasons for such results. The findings also show that the researched students on average studied only 40 minutes a week. This study suggests offering more courses with active learning approaches to intrinsically motivate students in order for them to spare more time for learning.

Keywords: higher education; learning assessment; PROG

Introduction

Assessment of higher education learning has been considered increasingly important (Clouder et al., 2012; Kushimoto, 2010; Sambell et al., 2012). Yet, measuring how much students learn has been a challenge for stakeholders involved in higher education assessment (Hardison and Vilamovska, 2009). While the grade point average (GPA) has traditionally been used for measuring students' academic performance at university, it is not considered a reliable indicator for learning because grading varies according to institutes, instructors, and other factors (Shavelson 2009). For instance, "As" in one institute and in another institute (or from one instructor and another instructor) may not have the same academic value. The implication of GPA from a certain year to another is also difficult to interpret, as they do not measure the same components. Even if a student's GPA was "A" in the first year and "B" in the second year, it is infeasible to determine that his or her learning deteriorated because the level of academic content may differ in the first year and the second year. In this context, various kinds of assessment tools have been created to measure university students' learning. This paper will review these assessment tools and then quantitatively and qualitatively analyze how much students improved generic skills at a Japanese university during the first two years of higher education.

Literature Review

Learning assessment tools

In the 1980s, tests such as the College Outcome Measures Program (COMP), the Academic Profile, and the Collegiate Assessment of Academic Proficiency were developed in order to measure the outcomes of general education programs. Especially, COMP drew attention from stakeholders in higher education assessment because it evaluated value-added learning: how much students learned during a certain period of time. COMP has two different forms: the Objective Test (multiple-choice questions) and the Composite Examination (multiple-choice questions, essays, and speeches). The Objective Test comprises of multiple-choice questions and the Composite Examination comprises of multiple-choice questions, essays, and speeches. Despite its popularity, COMP has been criticized for its inability to measure communication skills as well as critical thinking skills.

In the late 1980s, the College Basic Academic Subjects Examination (College BASE) was introduced. Apart from the subject content areas (i.e., English, mathematics, science, social studies), the College BASE assesses generic skills: interpretive reasoning, strategic reasoning, and adaptive reasoning. The College BASE has three forms: 1) the long form with content areas, 2) the short form with English and mathematics, and 3) an institutional-matrix form. As is the case with

COMP, however, a study revealed that the test assesses only a fraction of generic skills (Pike, 2011).

It is in this context that CLA has emerged as one of the most popular assessment tools of generic skills in higher education in the United States (Klein et al., 2007). Other parts of the world now recognize CLA because the Organization for Economic Cooperation and Development (OECD) has been developing Assessment of Higher Education Learning Outcome (AHELO) based on CLA (Douglas et al., 2012). CLA is an open-ended, value-added, performance assessment tool that measures generic skills such as critical thinking, analytical reasoning, problem solving, and written communication through writing tasks, make-an-argument tasks, critique-an-argument tasks, and realistic performance tasks (Council for Aid to Education, 2013).

In the United States, Arum and Roksa (2011) examined the CLA scores of 2,322 students at 24 universities over two years between the beginning of their first year in 2005 and the end of their second year in 2007. The study indicates that undergraduate students on average improved generic skills by 7%. While there are no universal standards for learning in higher education, they argue, students' gains in academic performance were low. They concluded that the poor result was attributed to the fact that the US college students on average study only 12 hours a week.

Studying hours of university students in the US and Japan

According to the National Survey of Student Engagement (2011), even 20 hours of studying time per week is not sufficient to fully prepare students for class. Apparently, both instructors and students have "If you don't bother me, I won't bother you" attitudes. That is, Kuh (2003) explains, instructors do not make students work hard so that they do not have to grade many papers or exams. Peters (2011) also argues that university instructors are responsible for students' low study habits. While students make as little effort as possible for their course work, university instructors tend to minimize course preparation time. The situation, however, may be even worse in Japanese higher education as Japanese university students study much less than their peers in the United States. As shown in Table 1, nearly 70% of Japanese university students study less than five hours a week.

	Japanese university students	US university students
0 hours	9.7%	0.3%
1-5 hours	57.1%	15.3%
6-10 hours	18.4%	26%
11-15 hours	7.3%	22.3%
16-20 hours	3.2%	16.8%
21-25 hours	1.9%	9.8%
26- hours	2.4%	9.5%

Table 1. Studying hours per week

University Management and Policy Research Center/National Survey of Student Engagement (Tsuji, 2013)

Tsuji (2013) estimates that the average number of hours spent studying by Japanese university students is only 3.5 hours. He explains that while Japanese university students are no less intelligent than students in other nations, their analytical reasoning and problem-solving skills are not well developed due to their lack of studying time. MacVeigh (2002) echoes Tsuji, stating that many Japanese university students are unable to think and write critically and logically and describes Japanese higher education as "a nationwide educational failure" (p.4).

Tsuji (2013) argues that there exists what he calls "(p. 77) spiral" between Japanese industry and higher education: companies' human resources personnel believe that university students do not study and they dedicate themselves to their part time jobs and/or circles/clubs, but companies' human resources personnel want students to study to acquire generic skills that lead to employability. University students claim that human resources personnel do not consider GPAs for job applications. Students thus spend more time on part time jobs and circles/clubs because they believe that it is more important for their future employability. Instructors are afraid that if they made students study hard, students would evaluate them poorly in the course evaluations. It is therefore better for instructors to give "whatever" lectures without sufficient preparation and spend more time on their research. Students then complain that instructors do not teach them well and spend even more time on part time jobs and circles/clubs instead of studying.

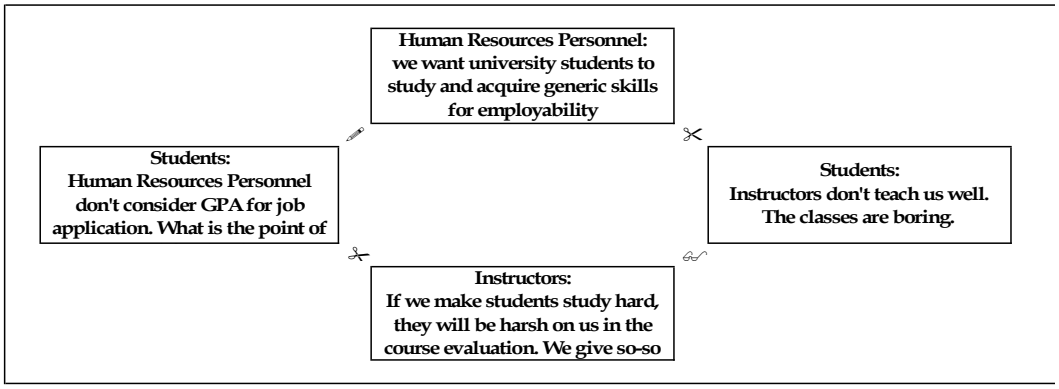

Figure 1. Negative Spiral (Tsuji, 2013)

In 2011, Tsuji (2013), as part of his NPO project, conducted surveys with 2,000 senior students in 28 departments at nine prestigious Japanese universities such as Waseda, Keio, Hitotsubashi, and Sophia. The student participants reported that only four out of approximate forty courses that they took at university helped them learn to think. A student, for example, reported that "Instructors didn't ask us any questions. At the end of the semester, we were only given a one-page report for evaluation." Another student claimed that "Professors just read textbooks in front of us." Tsuji's study indicates that Japanese university students are not in an environment where they are encouraged to study and develop their generic and employability skills. However, how can we know how much students learn at university in Japan? One answer is to employ learning assessment tools such as CLA and PROG.

PROG

This study employs PROG in an attempt to measure students learning. PROG examines two sets of generic skills: *literacy* and *competency*. This usage of the terms *literacy* and *competency* can be confusing as the elements of literacy and competency overlap. For instance, the Organization for Economic Cooperation and Development (OECD)'s definition of literacy—using tools interactively (e.g. language, technology)—is one of its Definition and Selection of Competencies (DeSeCo)'s key competencies. As Matsushita (2010) puts it, these tools include non-cognitive elements such as social and emotional elements that are part of *competency* (See Figure 2).

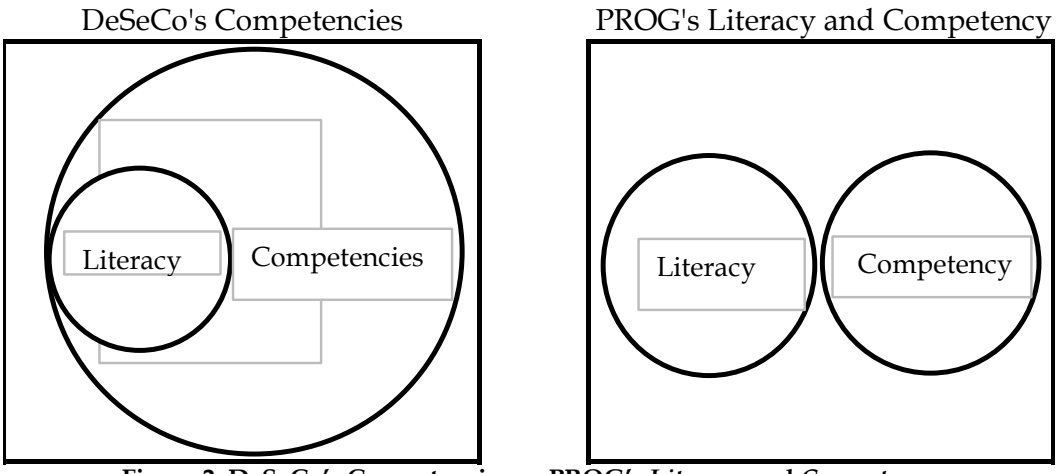

Figure 2. DeSeCo's Competencies vs PROG's *Literacy* and *Competency*
Based on (Matsushita, 2012)

In Japan, however, *literacy* is separated from *competency* (Matsushita, 2010). Despite weaknesses in the terminology, this paper follows PROG's usage of the terms *literacy* and *competency* but uses them in *italics* to differentiate them from OECD's usage.

In the PROG test, 45 minutes are allocated for the *Literacy* section and 40 minutes for the *competency* section. *Literacy* is composed of data collection, data analysis, problem solving, and conceptual thinking skills. Critical thinking skills, which are considered important generic skills, are partially integrated into the data analysis skills. PROG's *literacy* assessment also involves a few short essays to measure written communication and other skills. These elements are similar to what CLA examines. *Competency* is composed of skills in general communication, collaboration, networking, leadership, negotiation, and stress management as well as problem solving that is also included in *literacy*. According to Kawaijuku and Riasec (2015), problem-solving skills in *literacy* are tested on whether students can solve problems logically while problem-solving skills in *competency* are tested to what extent students solve problems as young professionals do (as described later). There are some notable differences between CLA and PROG. CLA is composed of open-ended essays while PROG is based on a combination of short essays and multiple-choice questions. Also, CLA is designed to produce results at the institutional level such as school average scores while PROG is designed to produce results at the individual level, which is scores for each individual who takes the exam. PROG also provides feedback sheets after the test with suggestions of how to improve generic skills.

The PROG score ranges from 1 to 7 for both *literacy* and *competency*. Score 4 is the level desired to be reached by the end of the first year of university. Students with this score are expected to be able to adequately understand and rephrase information from documents and graphs. Score 7 is the level desired to be reached by the time of graduation. Students with this score are expected to be

able to organize data and demonstrate information derived from the data in academic writing and graphs. Students at this level are able to establish arguments logically (Riasec, 2012).

The following problem is an example of a question that intends to test conceptual thinking skills in PROG's *literacy* section (Riasec, 2012).

> You are going to do a group presentation about globalization for a university seminar. Each group has 10 minutes, including questions and comments. Each group can decide the aspects of globalization on which to focus. You have 20 days until the presentation. Please make a plan, showing the process of how you would proceed with the preparation, taking the following elements into consideration:
>
> a. Preparing a presentation
> b. Collecting information and selecting ideas
> c. Deciding a group theme
> d. Analyzing information
> e. Deciding the content
> f. Practicing and modification
> g. Reviewing the presentation
> h. Deciding the roles

Students are expected to draw a flowchart that illustrates what they would do and when they would do it (Riasec, 2012). PROG has a few short essay questions that are similar to CLA's make-an-argument prompt. For instance, students are asked to read questions and answer in writing. The following is an example of such as question.

> A university student, who travelled to South Korea the other day, said that while young Koreans did not understand Japanese, those who were close to 80 years old whom s/he met were fluent in Japanese. Why do you think elderly Koreans are able to speak Japanese fluently? Briefly write down the reason(s).

While a student could write a creative story, for example, that the Koreans had lived in Japan in their youth and learned Japanese, they are expected to write a short essay based on his/her knowledge of Japanese colonial education in Korea between 1910 and 1945.

How does PROG address *competency*? The following question is an example from the *competency* section (Riasec, 2012).

> You are a project leader of developing a new product at a company. One of your subordinates came up with an interesting idea about a new

product. However, it is difficult to develop the idea into a product. In order to do so, it is necessary to deal with X, a major company with a marketing network, and Y, a venture enterprise with innovative technology. However, your company has not dealt with them in the past. What would you do? Choose one of the following options.

A. I would tell my subordinate that it is difficult to develop his idea into a product though his idea is interesting
B. I would encourage my subordinate to think of a realistic method of developing his idea into the product
C. I would tell my subordinate that we would keep the idea and wait for an opportunity to arise
D. I would think of how to communicate with X and Y through acquaintances
E. I would contact Y immediately to see if we could collaborate with each other to develop the idea into a product

Arguably, there is no right or wrong answer for this question. How does PROG score *competency* then? Kawaijuku and Riasec (2013) explain that they administered the test to the young business leaders that are currently active in society and collected sample data. They then analyzed the patterns of this group's answers to each question. That is, PROG attempts to measure students' *competency* by comparing their answers with the young professional leaders' answers and how similar students' answers are to those of the young leaders. Similar answers to young leaders' answers score higher in *competency* while dissimilar answers score lower.

According to Kawaijuku and Riasec (2014), as of April 2014, more than 100,000 university students took PROG. The average scores of PROG are 3.89 in *literacy* and 3.22 in *competency*. Approximately 63% of the test takers were first year students, 13.3% second year students, 19.7% third year students, and 3% fourth year students.

Year	*Literacy*	*Competency*
First year students	3.82	3.20
Second year students	3.97	3.16
Third year students	3.97	3.23
Fourth year students	4.28	3.58
Total	3.89	3.22

Table 2. PROG Scores

As Table 1 shows, students' scores in *literacy* improve through the entire course of university; however, the results show that there exists no improvement from the second to third year and then a remarkable improvement from the third to the fourth year. In *competency*, scores deteriorated slightly from the second to third year but improved from the third to fourth year. Although the number of

test takers is smaller than in other years, fourth year students substantially improved in both *literacy* and *competency*. It may be important to note, however, that Japanese fourth year students hardly go to university due to job hunting. Their improvement in generic skills, therefore, may not be attributed to university education, as described in the next section.

NUCB Education: Developing generic skills toward learning goals

The current research examines the learning of the first two years of education at the Nagoya University of Commerce and Business (NUCB). This university has participated in a national project entitled Improving Higher Education for Industrial Needs. In the context of improving university students' employability to meet industrial needs, NUCB started focusing on generic skills and thus employing PROG to assess students' generic skills.

Given that this study examines how much students improved generic skills in the first two years of NUCB education, it may be important to explain the first year experience (FYE) program called the Vision Planning Seminar (VPS) and the following year seminars at this university. The purpose of VPS is to help the first year students acquire generic skills and envision the professional careers based on the assumption that if they can envision their futures in the early stage of university life, they should be able to set goals and work toward acquiring skills necessary to achieve those goals. As shown in the Table 1, students are explicitly expected to acquire generic skills such as critical thinking, analytical reasoning, problem-solving and writing.

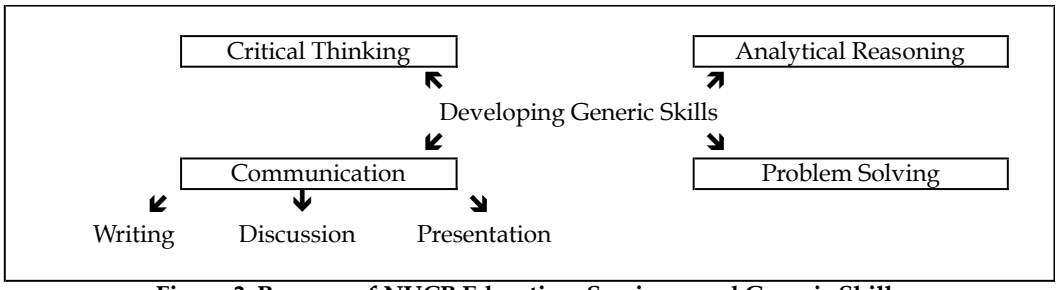

Figure 3. Purpose of NUCB Education: Seminars and Generic Skills
The table constructed based on the skills measured by CLA (Council for Aid to Education, 2013)

The generic skills acquired in VPS feed into the second year to fourth year seminars and for writing the bachelor's thesis. NUCB has set eight learning goals (LGs) to be achieved before graduation. LGs are assessed through the following skills:

1. Establishing a concrete, important, and feasible research theme
2. Acquiring academic knowledge and applying it to solve problems
3. Arguing convincing conclusions through proper processes
4. Expressing arguments through organized structure in writing
5. Creating and performing presentations

6. Communicating messages clearly and effectively
7. Utilizing information technology
8. Thinking and acting globally

Students are expected to demonstrate the first four of these eight skills in their bachelor's thesis and the last four in their second to fourth year seminars. These skills are generic skills as well. For instance, the skills to set a research topic or apply knowledge entail critical thinking, analytical reasoning, problem solving and writing. After all, generic skills overlap with basic research skills in many respects.

Methodology

The current research examines the PROG scores of 45 NUCB students who took PROG tests twice, first in April or May 2013 and second in December 2014, and analyzes the difference between these scores. That is, this study focuses on how much NUCB students learn in the first two years of university. One of the limitations of this study is its low sample size, which may hinder the generalization of the result. Yet, this type of value-added assessment is still rare in Japan and serves as an exploratory study. Some may also point out that the research period of two years is insufficient for this type of longitudinal value-added study. As Arum and Roksa (2011) affirm, however, "most of the gains in generic skills occur in the first two years of college…seniors do not spend much more time studying than freshmen" (p. 36-37). Although Kawaijuku and Riasec's (2014) study shows that Japanese university students substantially improved generic skills in the fourth year, the majority of Japanese university students do not attend regular courses in their fourth year and thus their improvement in generics skills are more likely attributed to their own study for the Synthetic Personality Inventory (SPI), an aptitude test for the selection of personnel, or related exams to seek employment, not university course work. Thus, the first two years of university learning is a reasonable indicator for the overall learning at university.

Signed Wilcoxon's rank sum test was conducted to examine whether there was a significant difference in the medians of the PROG scores between April/May 2013 and December 2014 ($p<0.05$) In selecting a statistical hypothesis test, we conducted the Lilliefors test based on the Kolmogorov-Smirnov test, which can determine whether sample data is normally distributed and found that p-value was less than 0.05 for all sample data, which shows that the sample data do not seem to be normally distributed. We thus selected Wilcoxon's rank sum over matched sample t-test that examines the means of sample data difference.

The analysis of PROG scores is supplemented by interviews with students who took PROG. The interview questions explore the following issues: how many hours a week students study how many hours a week students work and how

they have found courses at NUCB. The result of the current research might also be supplemented in the future by another longitudinal research of the same students at the end of their fourth year.

Results

	N	Means	SD	Min.	Max.
L1	45	3.09	1.427	1	6
C1	45	2.91	1.379	1	6
L2	45	3.16	2.225	1	7
C2	45	3.20	1.375	1	6

Table 3. Signed Wilcoxon's rank sum test: Descriptive statistics

*L1: *Literacy* score in the first year; L2: *Literacy* score in the second year

*C1: *Competency* score in the first year; C2: *Competency* score in the second year

While students' PROG scores improved by 0.07 (2.27%) in *literacy* and improved by 0.29 (9.97%) in *competency*, the difference of the medians between L1 and L2 was not statistically significant at any critical value (p-value=0.394). The difference of the medians of the medians between C1 and C2 was not statistically significant at $p<0.05$ but significant at $p<0.10$ (p-value=0.057). ~~Given that Arum and Roksa (2011) consider 7% improvement in CLA for two years "not much" in the US context, the PROG results of Japanese students can also be interpreted as "not much."~~

		N	Mean Rank	Sum of Ranks
L2-L1	Negative	17[a]	16.59	282.00
Ranks		17[b]	18.41	313.00
	Positive	11[c]		
Ranks		45		
	Ties			
Total				
C2-C1	Negative	13[d]	15.23	198.00
Ranks		20[e]	18.15	363.00
	Positive	12[f]		
Ranks		45		
	Ties			
Total				

Table 4. Signed Wilcoxon's rank sum test: Rank

a. L2<L1, b. L2>L1, c. L2=L1, d. C2<C1, e. C2>C1, f. C2=C1

In *literacy*, 17 students scored worse in the second test than in the first test while 17 students improved their scores. In *competency*, 20 students improved their score while 12 worsened it. Overall, NUCB students little improved generic skills for the first two years of higher education, though the results are not statistically significant and further studies at a larger scale are required to generalize the results.

With regard to interview results, while on average students spent 13.5 hours working, they spent only 40 minutes studying per week. The studying hours of NUCB students mark below the average studying hours of Japanese university students, which are 3.5 hours, as indicated in the literature review section. The maximum studying hours of NUCB students per week were three hours while one third of the students reported that they did not study at all. Three-quarters of students reported that they have encountered courses that they found interesting such as marketing, management and statistics. However, the number of courses that students found interesting is limited to a few out of many others. Some students claimed: "There is not much difference in content between high schools and universities," "The university courses are boring," "I always sleep in class," "High school teachers teach better than university professors," "One class has too many students," and so forth. One student also lamented that she had never received any feedback from professors regarding her assignments.

Discussion

This study indicates that in terms of PROG scores, NUCB students little improved generic skills during the first two years of higher education. *Literacy* did not improve at any significance level. While it is possible that NUCB students have made a statistically significant improvement in *competency* at $p<0.10$, though not at $p<0.05$, they might have acquired *competency*-related skills through interactions with others off campus (e.g., part-time jobs).

Apart from descriptive interpretations of PROG results, there may also exist two possible reasons/interpretations of the result: 1) PROG may not measure generic skills that NUCB intends to develop (or NUCB may not develop generic skills that PROG measures) and 2) (some) students make different levels of effort during the first and the second PROG tests.

Regarding the issue of incompatibility of measuring and developing different generic skills between PROG and NUCB, PROG, for example, does not measure ethical standards or global perspectives (Akihiro Tanabe, executive operating officer of Riasec, personal communication, February 18, 2015), both of which are nowadays considered generic skills and are elements of NUCB mission components. At the same time, while NUCB has set LGs to improve generic skills and measure them through students' self-evaluation and rubrics developed by NUCB's Assurance of Learning Committee, the university has been unable to provide evidence that it has developed generic skills comprehensively and objectively, as PROG results partially indicate. With regard to a plausible chance that some students make different levels of effort at the first and the second PROG tests, given that PROG results do not reflect on their GPAs or anything else, students may not motivated to try their hardest

when taking the PROG test. At any event, however, the sampled NUCB students study little as the study shows.

NUCB makes institutional efforts to increase students' studying time by establishing the self-study extra (SSE) point system. This system means that if students hours studying in the library, they earn academic benefits. If they spend 100 hours studying, for example, they can improve their grades by one level in five subjects, for example, from B to A. NUCB also sets strict standards for term-exams (one-fourth to one-third of test takers fail and must take re-exams to pass the course). Yet, NUCB students do not study much, as the surveys suggest.

Offering more courses that interest students is essential to intrinsically motivate them to learn. Employing active learning approaches may be recommended. As some scholars argue, active learning approaches help students learn to acquire generic skills (Barkley, 2010; Harper & Quaye, 2009; Ito, 2014c). A study examining an NUCB course employing active learning approaches also suggests that it has contributed to improving generic skills such as: agility, adaptability, and initiative; problem-solving skills; and curiosity and imagination (Ito, 2015).

This paper does not suggest that NUCB should completely redesign its curriculum to focus on improving generic skills that PROG measures; nevertheless, the university needs to better demonstrate how generic skills, which it intends to develop as expressed in its mission statement and LGs, are nurtured. Making and using rubrics to measure these skills is one option. NUCB's AOL has indeed developed rubrics to measure skills to achieve learning goals, though the committee struggles with implementation because rubrics can require substantial amount of time and effort to make and use. Administering a test that measures NUCB generic skills (e.g., global perspective) is another option. Riasec, the company that administers PROG, has flexibly dealt with measuring generic skills that PROG cannot measure. In order to measure students' global perspectives, for example, the company has engaged 735 professionals (aged between 25 and 49), who worked for global companies and managed foreign subordinates, to take the *competency* portion of the exam. These data can be used to make comparisons with students to know how similar these students are with global professionals: in other words, how global their perspectives are (Kawaijuku & Riasec, 2014).

Although this paper discusses the case of a particular Japanese university, the issue of developing and measuring generic skills are applicable to other contexts. Any university has to identify generic skills needed by their students and its institutional role in promoting generic skills. It then needs to develop or implement an effective set of tools to measure students' generic skills.

Conclusion

This study quantitatively examined how much students improved generic skills at a Japanese university during the first two years of higher education and qualitatively explored possible reasons for such results. The findings show that Japanese university students did not improve their generic skills by much during the first two years of higher education, arguably because students study little during this time. This study suggests that offering more courses with active learning approaches to intrinsically motivate students to spare more time for learning can contribute to improving generic skills.

Acknowledgement

We are indebted to Mr. Hirotaka Nishio at NUCB for his support in this research. We are also grateful to Mr. Ezra Anton Greene at University of British Columbia for editing this paper.

References

Arum, R. and Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago and London.

Barkley, E. F. (2010). *Student engagement technique: A handbook for college faculty*. New York: Jossey-Bass

Clouder, L., Broughan, C., Jewell, S. & Steventon, G. (2012). *Improving Student Engagement and Development through Assessment.* London and New York: Routledge.

Douglas, J., Thomson, G. & Zhao, C. (2012). The Learning Outcomes Race: the Value of Self-Reported Gains in Large Research Universities. *Higher Education* 64(3): 317-335.

Hardison, C. M. and Vilamovska, A. (2009). *The Collegiate Learning Assessment: Setting standards for performance at a college or university*. Santa Monica: Rand Education.

Ito, H. (2014a). Assessing an assessment tool of higher education: The case of PROG in Japan. *International Journal of Evaluation and Research in Education 3*(1): 1-10.

Ito, H. (2014b). What's wrong with learning for the Exam? An assessment-based approach for student engagement. *Journal of Education and Learning 3*(2): 145-152.

Ito, H. (2014c). Shaping the First-Year Experience: Assessment of the Vision Planning Seminar at the Nagoya University of Commerce and Business. *International Journal of Higher Education 3*(3): 1-9.

Kawaijuku and Riasec. (2013). *PROG: Progress Report on Generic Skills*. Tokyo: Kawaijuku and Riasec.

Kuh, G. D. (2012). What we are learning about student engagement. *Change* 35: 28.

Kushimoto, T. (2012). Outcome Assessment and its role in self-reviews of undergraduate education: in the context of Japanese Higher Education Reforms Since 1990s. *Higher Education* 59: 589-598.

McIntyre, J., Todd, N., Huijser, H., and Tehan, G. (2012). Building pathways to academic success. A Practical Report. *The International Journal of the First Year in Higher Education 3*(1): 109-118.

McVeigh, B. J. (2002). *Japanese higher education as myth*. New York and London: An East Gate Book.

National Survey of Student Engagement. (2011). How much time college students spend studying varies by major and corresponds to faculty expectations, survey finds.

Available at: http://nsse.iub.edu/NSSE_2011_Results/pdf/NSSE_2011_Press_Release.pdf.
Peters, R. A. (2011). Enhancing academic achievement by identifying and minimizing the impediments to active learning. *Public Administration Quarterly 35*(4): 466-493.
Pike, G. R. (2011). Assessing the Generic Outcomes of College: Selections from assessment measures. San Francisco: Jossey-Bass.
Riasec. (2012). *Measuring Generic Skills*. Tokyo: Riasec.
Sambell, K., McDowell, L. and Montgomery, C. (2012). *Assessment for Learning in Higher Education*. London and New York: Routledge.
Shavelson, R. (2009). *Measuring College Learning Responsibility: Accountability in a New Era*. Palo Alto: Stanford University Press.
Tsuji, T. (2013). *Nagze nihon no daigakusei wa sekai de ichiban benkyo shinai no ka* [Why Japanese university students study the least in the world]. Tokyo: Toyo keizai shinbunsya.