THE LIGHT AND SHADOW OF ASSESSMENT PRACTICES: FOCUSING ON SGH IN JAPAN

Kohei Nishizuka, Baozhu Wang, & Masahiro Arimoto
Graduate School of Education, Tohoku University (Japan)

Abstract

This research will discuss the assessment practices in global education in the context of upper secondary education in Japan. It will also aim to clarify some of the features, challenges and measures of methods for assessing learning outcomes, progress and transformation. In 2014, MEXT started a Super Global High School program, a national project toward fostering global human resources. A whole-school approach is utilized to design curriculum, such that teachers put an emphasis on global consciousness, and encourage the connection between local and global aspects by the students themselves. This leads to inquiry-based learning that inevitably results in the acquisition of competencies, and expands learning resources beyond classrooms. A comprehensive quantitative and qualitative analysis based on practical reports for each fiscal year and students' works that each SGH provided, and interview and questionnaire surveys to teachers and students was utilized. Some features, challenges and partial measures for formatively assessing student-centered inquiry learning that differs from didactic education types, will be revealed in reference to the diagnostic tool of curriculum evaluation that was originally developed based on the critical comments of MEXT's interim evaluation. The actual situation and issues to overcome regarding the formative and summative function of assessment and their linkage to global education in Japan, very relevant to the SDGs, will be considered in the discussion part to show the light and shadow.

Keywords: Super global high school (SGH), inquiry-based learning, assessment practices, formative assessment, whole-school approach.

1. Introduction

To "foster globalized leaders who will be able to play active roles on the international stage through education at high schools that contribute to this mission," Japan began a Super Global High School (hereinafter, SGH) Program in 2016 (Super Global High School, n.d.). SGH practices, which have been implemented as part of global education, have the same concepts as the ESD (Education for Sustainable Development) in the SDGs 4.7 (Sustainable Development Goals 4.7), which is associated with the SBCD (School Based Curriculum Development) that was developed to respond to the societal demands to foster global human resources and actualize individually optimized education. In addition to its four pillars of education, UNESCO also suggested "learning to transform oneself and society" as a new pillar to both explicitly and implicitly embody the aforementioned philosophy (UNESCO, 2009). Since the Tohoku Earthquake and Tsunami in 2011, the OECD Education 2030 Project has been cooperating in SGH's active learning approaches through policy dialogs with Japan (Arimoto et al., 2017).

However, implementing 21st century education models and assessment systems that include global learning methods such as inquiry-based learning and problem-based learning is a global challenge as balancing accountable and empirical assessment is extremely difficult because of the many contextual influences (Hargreaves, 2005; James, 2008; OECD, 2005). In accordance with the OECD (2005), this paper distinguishes assessment (judgments between the teaching and the learning), and evaluation (judgments between the teaching and the curriculum).

Of all the issues embodied in the ESD concept, the assessment for learning improvements is only the tip of the iceberg. Therefore, this research targeted the global education assessment practices in an SGH context to clarify the features, challenges, and partial measures needed to formatively assess student-centered inquiry learning, explain the differences from didactic education assessment, and provide direction on the development of appropriate assessment methods for global education practice.

The development of the SGH program, which gives designated schools financial assistance for five years, has been ongoing for more than five years. Initially, during the first phase, from the announcement of the SGH concept up to the public call for schools to participate in FY2014, the Minister of Education, Culture, Sports, Science and Technology (hereinafter, MEXT) in Japan was focusing on improving English language proficiency and globalizing Humanities and Social Science education. The initiatives for English language education improvements were actively adopted; however, in the second phase, which overlapped with the project start, the initial focus on the Humanities and Social Sciences was weakened and was eventually replaced by a problem-based learning approach focused on the Natural Sciences (Arimoto et al., 2017). As a result of the evaluation of the initial SGH outcomes, in FY2015, the program was again redefined and a basic foundation established in the third phase. From FY2017, there has been data available on curriculum development and practices, and the project is now focused on the fourth phase of stabilizing the educational activities.

2. Methodology

At the beginning of August, 2018, surveys were mailed to all 123 faculty members in charge of the SGH projects along with requests for the schools' annual research, development reports, and copies of student work. The survey had 26 Likert-type questions to be rated based on a school consensus as well as three open-ended questions focused on the efforts made by the school and their associated practices. By December 31, 2018, 55 schools had responded, a return rate of 44.7%.

Mid-project evaluations were conducted in the third year on all 123 SGHs to evaluate their progress and provide remedial action if necessary, all of which were completed in 2018. Although to date no school has withdrawn, withdrawal notices are to be given to schools in which there have been no improvements. Their practices are to be rated from one to five and each school will receive three comments regarding their characteristics, strong points, and shortcomings. While the evaluative criteria have yet to be made public, the first author used open coding to divide the comments into 22 broad classifications, and a checklist was developed outlining the identified good SGH practices. The analysis in the following section was based on the reports and the survey results for the following items: C05 (Effective Use of the Administrative Guidance Committee), C07 (Organization Management where All Teachers are United), C16 (Process-oriented Inquiry Activities), C18 (Curriculum Evaluation), C19 (Assessment of Learning), and C22 (Student Agency).

3. Results and discussion

To deepen learning and enhance the quality of the school's activity systems, assessments needed to be formative. If there was a lack of transparency and organization, the assessment methods used to assess the learning progress and adjust and expand the learning would possibly only reflect a fraction of the overall learning process. Therefore, understanding the ambiguities related to the formative assessment (hereinafter, FA) and knowing how to deal with the possible barriers was vital to revealing the various SBCD inquiry-based learning assessment characteristics and implementing strategies with certainty. Five issues, I01 to 05, were extracted, that revealed the fragmented practices at different SGHs and offered insights into the assessment issues, as discussed in more detail in the following.

I01: Teachers do not have sufficient time to get involved in feedback and feedforward for students.

Teachers found it difficult to find the time to fully assess the inquiry activities of the individuals or small groups one by one as they had to do their everyday work as well as manage the SGH. From the survey, 46 of the 55 schools were worried about the teachers' workload. Some schools had external partners such as university and graduate students and were able to take advantage of an ICT system for remote feedback and feedforward for the students. However, as such external partners require financial support, the schools will have to cover these expenses once the designation period ends.

Another partial measure would be to dedicate time to interact with the students, as exemplified at High School A. When the first author conducted the interview research (February 1, 2019), the teacher in charge of the SGH project emphasized that he wanted to "build a school where everyone can continue learning." To achieve this, he and his colleagues had been attempting to develop a more flexible curriculum to break down the barriers between the subjects and planned to continue the inquiry activities after their designation ends in 2019. One example is that they plan to make use of one of the outcomes that has been cultivated as part of the SGH practices; that is, "classes will be limited to 6 periods, with teachers having the freedom to teach what they want for the 7th period." This freedom does not mean another class based on the interests or specialties of the teachers, but a time reserved for student

interaction to optimize individual student-teacher learning. The school is seeking to respond to the student needs by establishing a flexible curriculum that honors the students' decisions.

I02: The school assessment methodology is undeveloped.

Generally, to develop the SGH performance activities assessments, support is given by senior students *(senpai)* and university researchers. The SGH committee concurrently develops the curriculum, research themes, pedagogy, and assessment methods. However, as the development of the assessment tools is given a lesser priority than the development of the course content and the associated teaching and curriculum values, focus is placed on visualizing the learning outcomes and validating the summative assessment (hereinafter, SA) learning transformation. On the other hand, in some schools, the assessment methods were developed by devoted groups within the school organization to ensure the implementation of formative feedback. Therefore, it appeared that the focus on developing appropriate assessment methods depended on the school scale, the allocation of duties, and the availability of assessment methodology instruction from university teachers and specialists. For example, High School B established an international fieldwork research team, a domestic fieldwork research team, a publicity/information and research outcome/practice report team, an accounting team, an assessment research team, an inquiry course research team, and an English class inquiry team as a part of the SGH promotion committee.

However, it should be noted that what was mentioned above means teachers in Japan just don't function the Western-oriented theory of assessment. As Arimoto and Clark (2018) pinpointed the cultural identity on which the cognitive assessment has the bedrock, assessment events mediated by cultural scripts can be seen with the unorganized, improvised, and informal traits. They might be called Japanese-style assessment to achieve the responsibility that builds loose interrelationships between teachers and students, which is perfectly independent from the accountability.

I03: Feedback functionality depends on psychometrics assessments.

If FA was not adequately conducted at the school, standardized tests developed by external private companies based on a psychometric paradigm were utilized to ensure accountability for the educational outcomes, reduce the labor burden, and allow for the visualization of the outcomes for authorities. As this type of assessment is merely a measure or a superficial snapshot of improvement, it runs the risk of being taken as a be-all assessment (Hargreaves, 2005). While many SGHs were attempting to develop tools that worked with the school assessment tools such as rubrics as well as using the objective tests, it was not clear whether the measures had clear and positive correlations (refer to I04).

The methodology being used to validate the SGH program course effectiveness was also questionable. It was found that some schools conducted consciousness surveys as indirect SA for which a comparison was made between students who had engaged in SGH inquiry activities and students who had not. For example, to assess the validity of the SGH learning activities at High School C, students who were not engaged in the SGH (the control group) were compared to the students who were engaged in the SGH (the experimental group), from which it was found that the engaged students had greater global competencies; however, in High School D, there were few statistically significant differences between the non-SGH and SGH students. In High School E, it was found that the scores for students who were partially engaged in the inquiry activities declined compared to non-SGH students. These results raise the question as to what the SGH curriculum brings to education and also cast doubt on the validity of SA such as self-assessment type consciousness surveys.

I04: The curriculum is evaluated based on summative de-contextual assessment.

In general, because the curriculum evaluations were designed in reference to the learning assessment data (OECD, 2005), schools tended to employ validation methodologies that measured the degree to which the students were influenced by the SGH curriculum. It should be noted that these methods were not assessments that reflected the students' learning situations, but rather were decontextualized assessments. Essentially, curriculum validation should be conducted by extracting the assessment data related to its related inquiry-based learning context; however, it was found that the curriculum was being validated based on information that had been separated from the actual learning activities; that is, decontextualized objective tests and general consciousness surveys.

Most SGHs have been using rubrics as the assessment tools for the inquiry activities. If the performance assessments related to the rubrics are unsatisfactory but the external objective test scores taken at the end of the term are high, is it justifiable to say that this was a result of SGH inquiry-based practices? There are also cases where students had high total rubric scores but poor standardized test scores. These results indicate that the decontextualized assessment data in many of SGH annual reports resulted in a misinterpretation of the effectiveness of the curriculum and a lack of reliability on the inquiry-based learning assessment procedures; that is, there was a disconnection between the situated inquiry-based learning assessments and the decontextualized assessments, which meant that either the curriculum organization or the inquiry activity assessments method were invalid or that the measurable abilities of the situated inquiry activity assessments differed from the decontextualized assessments.

For example, High School F conducted a cross analysis between the assessment scores and the standardized tests for the "thinking skills" from the private company and the teacher assessed rubric scores for the "presentation" or "research papers" activities, for which positive correlations were found in 2016 (second year of designation), but not in 2017 (third year). Therefore, it is necessary to further investigate the validity and benefits of comparing the assessments, and fully understand the different psychological constructs behind the school-based learning standards and the standardized tests.

At High Schools E and G, to assess the summative transformation, a text-mining approach was used to quantitatively analyze the qualitative data in each student's research paper. Qualitative and holistic judgment activities were included as a part of the third-generation assessment procedures, which were based on socio-cultural constructivist framework in which the assessments are embedded in the situated learning (James, 2008). Because these individually specific data are not enough evidence to prove overall curriculum effectiveness, the rubrics offer nothing more than supplementary information on the learning activity. Therefore, it is preferred that the comprehensive assessment data be re-contextualized at a group level in inquiry activity curricula. To ensure accountability, educational organizations are also interested in summative evidence of the decontextualized and factorized learning outcomes. Indeed, the fact that the terms assessment and evaluation are synonymous with "*hyoka*" in Japanese, complicates this issue. If curriculum evaluation is confused with learning assessment, the decontextualized assessments will not reflect the nuances in the learning and the assessment activities aimed at determining the learning progress will become progressively isolated and diluted (see Figure1).

Figure 1. Assessment System in SGH.
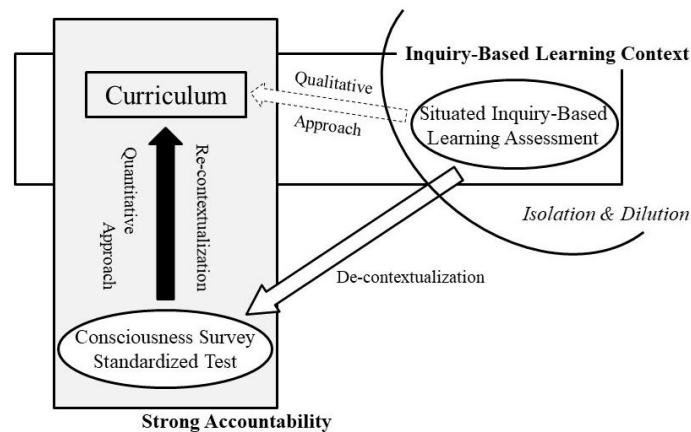
I05: Learning intentions and criteria are unilaterally decided by teachers.

This comment was not related to feedback and feedforward opportunities, but rather the quality. Most SGHs set performance issues every year, with the rubrics being developed, presented, and shared at the beginning of the fiscal year with the students. The standards and criteria have generally been based on the competencies that each school established as the learning intentions. However, what was found was a pseudo agency problem. Summative dominant-assessments based on the criteria were only developed by the teachers and were executed scientifically and technically without any consideration of the particular situation (Almqvist et al., 2017); that is, it appeared the aim of improving the students' learning was discharged and a top-to-bottom unequal power relationship retained between the students and teachers in which the students were only able to demonstrate free subjectivity within an instrumentalized learning situation. The students had constrained assessments under a pseudo freedom, which meant that the teachers were not associating with the students through FA, which in turn reduced the learning motivation of learners who were unable to keep up with the externally imposed instructions (Hughes, 2014).

It was found that many of the SGHs have been unable to shake off the passive attitude of "feeling forced." In reference to the student agency in the evaluation tools (C22), some schools said that they were having trouble with free riders in the group activities. Further, for the over ten SGHs that required external intervention by an administrative guidance committee (C05), it was reported that there were many students who had become demotivated in the presentations or problem-based learning research. This indicates that students reluctant to follow the teachers' instructions. Therefore, it may be necessary to develop situated and diverse assessments that are focused on each student's needs. Although this may be a time-consuming process, High School I had taken the time to develop criteria that specifically fit their students' personalities and strengths while at the same time embracing the teachers' qualitative and quantitative assessment aims. The value of the SGH project has been first and foremost to change the traditional secondary school focus on accountability and college entrance examination requirements, and to develop curriculum and assessment tools that are designed to enhance the students' global agency.

4. Conclusion

The research into the effectiveness of inquiry-based learning revealed the difficulties in adjusting from traditional subject-oriented learning. Inquiry-based learning allows students to get and give feedback and feedforward from teachers, peers and specialists, which lead to an expansion of their learning environment, give them the opportunity to participate in their learning, and develop a deeper dialog with the materials and a clearer understanding of their progress. What this research review indicated was that it has been difficult to fully understand student progress if the feedback and feedforward between the teacher and the student is insufficient and not focused on the student's specific learning style. Unfortunately, it appears that only a few teachers were bearing the burden, and there were few schools which took a whole-school approach in which all faculty and support staff work together to support the students; however, this is a prerequisite for a successful qualitative transformation. The "feedback" and "advice" received in the open-ended responses indicated that many teachers were unfamiliar with the educational philosophies underlying FA; however, As mentioned in I02, it is needed to deliberately observe the practices and school organizations, and to analyze social languages and narrative used in assessment events for the conceptualization of Japanese-style assessment.

Although there has been some FA research in Japan at the subject-oriented level, further discussions are necessary to address the issues found in our evaluation. One SGH had been able to resolve these problems, and even though their solutions may not be a cure-all for all schools, learning from good practice could give rise to further innovations through a retranslation through each community's cultural lenses (Phuong-Mai et al., 2005). The SGH curriculum and assessment are challenging issues for teachers who are not familiar with the theory. Nevertheless, by sharing and re-interpreting knowledge, each school will be able to transform and innovate while considering culturally appropriate strategies.

References

Almqvist, C. F., Vinge, J., Väkevä, L., & Zandén, O. (2017). Assessment as learning in music education: The risk of "criteria compliance" replacing "learning" in the Scandinavian countries. *Research Studies in Music Education*, 39(1), 3-18.

Arimoto, M., & Clark, I. (2018). Interactive Assessment: Cultural Perspectives and Practices in the Nexus of "Heart or Mind." In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge Handbook of Instructional Feedback* (pp.474-503). Cambridge: Cambridge University Press.

Arimoto, M., Nishizuka, K., Nomi, Y., & Ishimori, H. (2017). Proceedings of IAC in Vienna '17: *Pedagogical approaches to global education: follow-up study of TOHOKU School 2.0 since 2014* (pp.127-151). Prague: Czech Institute of Academic Education z.s.

Hargreaves, E. (2005). Assessment for Learning? Thinking outside of the (black) box. *Cambridge Journal of Education*, 35(2), 213-224.

Hughes, G. (2014). *Ipsative Assessment: Motivation through Marking Progress.* Basingstoke: Palgrave Macmillan.

James, M. (2008). Assessment and Learning. In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflections and application* (pp.20-36). Abingdon: Routledge.

OECD (2005). *Formative Assessment: Improving Learning in Secondary Classrooms.* Paris: OECD.

Phuong-Mai, N., Terlouw, C., & Pilot, A. (2005). Cooperative learning vs Confucian heritage culture's collectivism: confrontation to reveal some cultural conflicts and mismatch. *Asia Europe Journal*, 3(3), 403-419.

Super Global High School (n.d.). Outline of Super Global High School Program. Retrieved January 1, 2019 from www.sghc.jp/en/.

UNESCO (2009). Review of Contexts and Structures for Education for Sustainable Development 2009. Retrieved January 1, 2019 from http://www.unesco.org/education/justpublished_desd2009.pdf.