A Longitudinal Study of Effectiveness, Impact, and Challenges in the Japanese Quality Assurance System

Susumu Shibui *, Nozomi Takahashi *, Ayaka Noda *

Abstract

Higher education has seen a concern about improvement of the quality of external quality assurance. The Certified Evaluation and Accreditation (CEA), a Japanese quality assurance system, is currently in the middle stage of its second cycle and in discussion to improve its quality in the next cycle. In order to inform the development of a future system, this study aimed to explore how universities perceived the effectiveness, impact, and challenges of the CEA system and practices and how these changed between the first and the second CEA cycles. Questionnaire surveys of the evaluation methods and arrangements, process, effectiveness, and matters of CEA system were administered to 52 universities accredited by the National Institution for Academic Degrees and Quality Enhancement of Higher Education immediately after the evaluation in the first and the second cycle. The universities' responses in the two cycles were analyzed and compared. Statistical analyses revealed changes in their perception which related to "effectiveness in helping quality enhancement," "gaining public understanding and support," and "evaluation exhaustion (workload)." The implications of these changes for the growth of the CEA system were discussed.

Keywords: Japanese Accreditation System, Questionnaire survey, Quality Enhancement, Public understanding and support, Evaluation exhaustion.

1 Introduction

Higher education has dramatically changed over the last decades as the world has changed, with globalization, economic growth, knowledge-based societies, increased needs for accountability, and university massification. These changes have brought a growing demand for both internal and external quality assurance systems for higher education. Where external quality assurance is concerned, quality assurance institutions have attempted to examine their own performance and verify the appropriateness of their evaluation system to support quality assurance in higher education.

Japan has developed a university accreditation system called the Certified Evaluation and Accreditation (CEA), whose aim is to assure quality higher education. The CEA has been defined as follows; (1) "Certified" by the Minister of Education, Culture, Sports, Science and

* National Institution for Academic Degrees and Quality Enhancement of Higher Education, Tokyo, Japan

Technology (MEXT); (2) "Evaluation" to promote quality enhancement in education and re-search by quality assurance institutions, and (3) "Accreditation" to assess whether a university fulfills standards. Japanese accreditation is still new. It was only in 2004 that accreditation be-came a legal requirement defined by the *Basic Act on Education*. All 4-year universities must receive external accreditation every seven years by one of three official accreditation agencies: the Japan University Accreditation Association (JUAA), the National Institution for Academic Degrees and Quality Enhancement of Higher Education (NIAD-QE), and the Japan Institution for Higher Education Evaluation (JIHEE). The university accreditation process is similar across the three agencies. Based on institutions' self-evaluation, document analysis and site visits are conducted by peer review. After the interactions between institutions and accreditors, final evaluation reports are released to the public.

In response to the international trends of quality assurance systems and accountability, Japanese quality assurance institutions have revised their evaluation standards in the next cycle of the CEA by analyzing various resources such as past evaluation reports, public hearings, or questionnaire surveys for universities and evaluators [1][2][3]. Although the new standards and systems differ according to the quality assurance institution, they primarily place more focus on measuring learning outcomes, internal quality assurance systems for teaching and learning, and assisting universities for accountability through public infor-mation disclosure as well as reducing the evaluation workload [4][5][6]. In response to these revisions, several research programs have started to attempt to provide indicators and evi-dence of student learning outcomes by analyzing university evaluation reports to support CEA [7][8][9]. The new CEA system has been implemented and is currently in the middle stage of its second cycle.

Although there has been considerable discussion regarding whether and how the newly revised system would help to improve the previous CEA system, and how it could stimulate development in anticipation of the next cycle, little study has been conducted to address these questions. Therefore, it is important to employ scientific or statistical methods to un-derstand how the new CEA system can improve its effectiveness and appropriateness. By analyzing and comparing the two surveys that NIAD-QE conducted in the first cycle (2005-2011) and the second (2012-2014), we undertook a study to examine how universities perceived the CEA system and practices and how these have been changed between the two cycles.

Our study investigated the NIAD-QE questionnaire survey, targeting the subject universities immediately after they completed the accreditation. Every year, the survey asks the accredited universities about the evaluation methods and arrangements, processes, practices, and matters (see Table 2) that should need improvement or further discussion for continued development of the evaluation framework, so that understand the appropriateness of the evaluation can be de-termined [10]. The survey also includes samples of institutions' actions taken to address needed improvements identified in the evaluation reports, and samples of NIAD-QE's actions taken in response to the questionnaire results (e.g., updating of the standards for evaluation and accredi-tation, simplifying of the method of site visits, etc.). In the first cycle verification, the results of the survey were analyzed in terms of the three objectives of CEA —assuring quality, helping improvements and achieving accountability—and showed that universities perceived the effectiveness of the CEA for assuring quality and helping improvements. However, achieving accountability of universities, which remained an issue, as did reducing the evaluation [2]. Based on the results of the analyses of the questionnaire survey and evaluation repots in the first cycle, NIAD-QE revised the university evaluation standards in the second cycle of the CEA (Table 1).

However because the survey in the first cycle was conducted immediately after the CEA

system, the positive response to the two objectives seemed to refer to the immediate impact of CEA; thus considerations regarding a longitudinal CEA impact seemed to be unanswered. To address this question, in the second cycle, the survey added new questions concerning the out-comes from the previous CEA and its impacts, making reference to the objectives of the evaluation, such as quality assurance of educational and research activities, helping universities make improvements, and achieving public understanding and support (Question 9). We analyze whether there are differences between the response to the questions asking about the effectiveness of the new CEA results (Question 6-(2)) and the previous CEA (Question 9) in the second cycle.

The study we are reporting on in this paper aims to examine how universities perceived the effectiveness, impact, and challenges of the CEA system and practices and how these have changed between the first and the second cycles, in order to develop a new system looking ahead to the next cycle. To investigate these matters, we have used two approaches to analyze the results of survey. First, the data was compared between the first and the second cycles. Second, we analyzed whether there were differences between the responses to the question asking about the effectiveness of the new CEA (Question 6-(2)) and the previous CEA (Question 9). The results are discussed in terms of following three themes: "effectiveness in helping quality enhancement," "gaining public understanding and support," and "evaluation exhaustion (workload)." As mentioned earlier, these themes were considered challenges for the CEA. Especially for helping quality enhancement, [11] has pointed out that if the results of the CEA have not been applied to improve the quality of higher education, the CEA has just been evaluation for the sake of evaluation. Reference [12] has also suggested that almost all universities in the first cycle established the CEA goals by taking the gathered final evaluation reports, analyzing them and describing enormous quantities of materials and resources, but they did not go on to attempt to apply problem and knowledge obtained through the CEA process to improve their quality. Reference [13] has mentioned, given the results of a questionnaire survey for institutional researchers, that most universities set systems for university evaluation; making use of the results of evaluation for management within universities was under development. Because necessary actions in response to evaluation results depend upon the self-initiative of each institution in the current CEA system [14], it is unclear whether and how universities make efforts to improve their actions based on CEA results. However, recent research has reported that approximately 70% of all the 53 universities accredited by NIAD-QE applied feedback called "improvement needed" that was pointed out in the accreditation results of their quality improvement, suggesting that CEA results worked as external pressure and /or incentive for quality improvement [15]. The study we are reporting today concentrates on the perception of accredited universities of the ef-fectiveness, impact, and challenges of CEA system and attempts to seek a way of enhancing the CEA system.

Table 1: Standards of Institutional Certified Evaluation and Accreditation for Universities

First Cycle: FY2005–2011	Second Cycle: FY2012–2018
1. Mission of university	1. Mission of university
2. Education and research structure	2. Teaching and research structure (organizations)
3. Academic staff and education supporting staff	3. Academic staff and teaching support staff
4. Student admission	4. Student admission
5. Academic programs	5. Academic programs (contents and methods)
6. Effectiveness of institutional performance	6. Learning outcomes
7. Student support	7. Facilities and student support
8. Facilities	8. Internal quality assurance system of teaching and learning
9. Internal quality assurance system	9. Finance and management
10. Finance	10. Public information disclosure on teaching and learning
11. Management	

Source: *NIAD-UE's Certified Evaluation and Accreditation* [5]

2   Method

Fifty-five universities accredited by NIAD-QE in both cycles (the first, 2005-2011; the second, 2012-2014) were targeted in questionnaire surveys. The number of responses was 52 universities (46 National universities, 5 Municipal/Prefectural universities, 1 Private university) out of 55 universities, thus the survey response rate was 95 %. The questionnaire was constructed in 11 sections (Table 2). Sections 9 and 10 were newly added in the second cycle. Sections 7, 8, 11 were excluded from our analyses because of the open-ended questions. Each section contains ap-proximately three to twenty-five items. (Tables 3-10).

Table 2: Questionnaire of the second cycle (The original is in Japanese)

1.   Evaluation standards and viewpoints
2.   Evaluation methods and contents
    2.1   About self-evaluation
    2.2   Site Visits
    2.3   Statements of objection
3.   The workload and schedule of the evaluation
    3.1   the workload required for the evaluation
    3.2   the period of operations set by NIAD-QE
    3.3   the amount of effort required for the evaluation
    3.4   the evaluation schedule
4.   Orientation meetings and training sessions
5.   Evaluation results (evaluation report form)
6.   The effects and impacts from evaluation
    6.1   The effects and impacts from self-evaluation
    6.2   The effects and impacts from NIAD-QE's evaluation results
7.   The use of evaluation results   (open-ended questions)
8.   Regarding the implementation system for evaluation recommendations (open-ended questions)

9. The effects from and impacts from your last certified evaluation
10. NIAD-QE's certified evaluation process in comparison with previous evaluations
11. Other (open-ended questions)

3   Results of Comparative Analysis of the Questionnaire

Means and standard deviations were calculated for the data from both cycles (Tables 3-10). Universities' responses were analyzed from two approaches. First, the data was compared between the first and the second cycles. Second, we analyzed whether there were differences between responses to the question asking about the effectiveness of the new CEA (Question 6-2) and the previous CEA (Question 9). In our approaches, *t*-tests and chi-squared tests were performed. A paired-samples *t*-test was conducted to compare the response in first and second cycle in each question, which requested an answer on a 5-point scale. With regard to the questions where the requested answers used a 2-point scale, a chi-squared test of independence was calculated comparing the frequency of answer in the first and second cycles. In cases where the expected numbers are less than five under any condition, we calculated Fisher's exact test of independence instead of a chi-square test.

The statistical hypothesis testing from one of the methods we have referred to above, significant differences were observed in nine out of all the 87 questions. We will now discuss the re-sponse tendency for each cycle as well as the open-ended questions. To interpret the tendency of responses to questions using a 5 point-scale, the score 3.5 was regarded as the neutral response, following the "yes" bias reported from the pilot study [16].

3.1   Evaluation standards and viewpoints

The responses to the two questions showed differences depending on the cycle. Question 1-5 showed a significant trend (χ^2 (1) = 3.30, p = .07), and Question 1-6 showed a significant difference (χ^2 (1) = 15.7, p < .01) in each cycle. These questions concerned the difficulty of preparing the self-evaluation reports.

Other questions did not show the significant differences by cycle. For Question 1-1 (t (51) = .20, p = .84), 1-2 (t (51) = -.65, p = .52), and 1-4 (t (51) = .16, p = .88), there were positive responses in each cycle. For Question 1-3 (t (51) = .97, p = .34), there were neutral responses in each cycle.

Table3: Summary of descriptive statistics for Questions 1 between 2 cycles (*N*=52)

	First Cycle		Second Cycle	
Questions	*M*	*SD*	*M*	*SD*
1. The composition and contents of the evaluation standards and evaluation viewpoints were appropriate for quality assurance of your institution's educational and research activities.	4.06	.50	4.04	.44
2. The composition and contents of the evalua-	4.06	.50	4.12	.47

tion standards and evaluation viewpoints were appropriate for supporting your institution's educational and research activities.				
3. The composition and contents of the evaluation standards and evaluation viewpoints were appropriate for supporting public understanding and your institution's educational and research activities.	3.73	.71	3.62	.56
4. The establishment of the evaluation standards and evaluation viewpoints for educational activities were appropriate.	4.15	.63	4.13	,62
5. Some evaluation standards/viewpoints were difficult to self-evaluate. [+ a), c)]	1.56	.50	1.36	.48
6. There were duplicate or overlapping evaluation standards or viewpoints. [** b), c)]	1.57	.49	1.18	.39

[a)] $N=39$, [b)] $N=49$, [c)] *two-points scale*
Notes. [+] $p < .10$, * $p < .05$, ** $p < .01$

3.2 Evaluation standards and viewpoints

Subsection 2-(1) contains the questions about the self-evaluation. Only Question 2-(1)-6 in this subsection showed a significant difference in each cycle (t (51) = -3.55, $p < .01$). There was a significant increase in positive responses in the answer to the question about the adequacy for character limit.

Other questions did not show significant differences (2-(1)-1, t (51) = -.39, $p = .70$; 2-(1)-2, t (51) = .20, $p = .84$; 2-(1)-4, t (51) = 1.54, $p = .13$; 2-(1)-5, t (51) = 1.00, $p = .32$; 2-(1)-7, χ^2 (1) = .80, $p = .37$). All the answers in this subsection except for Question 2-(1)-2 showed positive response in both cycles, whereas Question 2-(1)-2 showed negative responses in both cycles. Note that the answer of Question 2-(1)-3 was excluded, as the scale changed from a 5-point scale to a 2-point scale in FY2008.

Subsection 2-(2) contains the questions about site visits. None of the questions showed a significant difference (2-(2)-1, t (51) = .42, $p = .32$; 2-(2)-2, t (51) = .17, $p = .86$; 2-(2)-3, t (50) = .00, $p = 1.00$; 2-(2)-7, t (51) = 1.66, $p = .10$; 2-(2)-8, t (51) = - .64, $p = .53$; 2-(2)-9, t (51) = 1.00, $p = .32$). Note that the questions have changed in 2-(2)-4, and new questions are added in 2-(2)-5 and 2-(2)-6. These three questions are excluded from the statistical analysis.

Subsection 2-(3) contains the questions about the process of the statement of objection. Neither of the two questions showed a significant difference (2-(3)-1, t (51) = 1.20, $p = .24$; 2-(3)-2, t (48) = 1.44, $p = .16$). Question 2-(3)-3 had few answers and was excluded from statistical analysis because this question was only applied to institutions that had done the statement of objection.

Table4: Summary of descriptive statistics for Questions 2 between 2 cycles ($N=52$)

	First Cycle		Second Cycle	
Questions	*M*	*SD*	*M*	*SD*
(1) About self-evaluation				
1. You were able to appropriately conduct the	4.06	.60	4.10	.53

self-evaluation following the evaluation standards and viewpoints.				
2. You were able to appropriately include materials for the self-evaluation form (s) because you already had them.	2.90	.95	2.87	.83
3. You were unsure about which materials to include with the self-evaluation form(s). [c]	1.00	.00	1.00	.00
4. You were able to complete an easily understandable self-evaluation form to support general public understanding of your instituion's general situation.	3.83	.73	3.60	.74
5. You were satisfied by the degree to which the self-evaluation form was completed.	3.90	.60	3.77	.70
6. The word limit on the self-evaluation form was large enough to allow a satisfactory self-evaluation. **	3.15	1.08	3.83	.87
7. When completing the self-evaluation form, your institution referred to a previously completed self-evaluation form from another university that had undergone NIAD-QE's certified evaluation. [a, c]	1.90	.30	1.84	.37

(2) Site Visits

1. The contents of the "Situation of Document Analysis" presented before site visits were appropriate.	4.04	.44	3.96	.48
2. The contents of the "Checkpoints During Site Visits" presented before the visits were appropriate.	3.94	.69	3.92	.43
3. The contents of the questions asked by NIAD-QE's evaluating supervisors during site visits were appropriate. (Exclude the administrative supervisor in all questions for this section.) [b]	3.96	.68	3.98	.64
7. During site visits, a shared understanding was reached with NIAD-QE's evaluating supervisors regarding the current state of educational and research activities.	4.15	.63	3.98	.67
8. During site visits, the number and composition of NIAD-QE's evaluating supervisors were appropriate.	3.98	.69	4.16	.67
9. During visits, you believed that NIAD-QE's evaluating supervisors received adequate training	3.90	.63	3.84	.70

(3) Statements of objection

1. The implementation methodology and scheduling of statements of objection were appropriate.	4.08	.62	3.96	.66
2. The inclusion of the "Content of and Response to Statements of Objection" in the evaluation report form was appropriate. [a]	4.12	.69	4.02	.77
3. NIAD-QE's response to your institution's statement of objection was appropriate.	4.25	.83	4.25	.83

[a] *N=49,* [b]*N=51,* [c] *two-points scale*
*Notes. ** p < .01*

3.3 The workload and schedule of the evaluation

Subsection 3-(1) contains questions about workload. Question 3-(1)-2 (t (51) = -3.74, $p < .01$), 3-(1)-3 (t (51) = -2.97, $p < .01$), 3-(1)-4 (t (51) = -2.53, $p < .05$) showed a significant difference in each cycle. These three questions show that the universities perceived a workload increase for preparing site visits. Questions 3-(1)-1 (t (51) = -.64, $p = .53$) and 3-(1)-5 (t (51) = .32, $p = .76$) did not show a significant difference. From the trend of the answers in the two cycles to the Questions 3-(1)-1, we can conclude that the universities' perception of workload for preparing the self-evaluation reports remained heavy over two cycles.

Subsection 3-(2) contains questions about the evaluation schedule of evaluation set by NIAD-QE. Each question included questions about the workload and adequacy of the evaluation schedule. Questions about the adequacy of the evaluation schedule did not show any difference (3-(2)-1, t (51) = 1.12, $p = .70$; 3-(2)-2, t (51) = .39, $p = .70$; 3-(2)-3, t (51) = .44, $p = .78$; 3-(2)-4, t (51) = .11, $p = .92$). From the trend of these answers, universities perceive that the evaluation schedule set by NIAD-QE was satisfactory.

Subsection 3-(3) includes questions about whether the workload for the evaluation worth the cost or not. None of the three answers showed a significant difference (3-(3)-1, t (51) = -1.14, $p = .26$; 3-(3)-2, t (51) = -.12, $p = .90$; 3-(3)-3, t (51) = -.89, $p = .38$). The universities' answers to the question about whether the workload was worth the cost were positive compared to the answers to the questions that asked only about the workload in each cycle.

Subsection 3-(4) includes questions about the schedule of CEA. Neither of the two questions showed a significant difference (3-(4)-1, χ^2 (1) = 1.70, $p = .19$; 3-(4)-2, χ^2 (1) = 0.32, $p = .57$). The trend of the answer showed the universities' positive perception of the CEA schedule set by NIAD-QE.

Reports of the open-ended questions showed the following trends. The comments most frequently reported concern the workload for collecting the data and evidence. One can also observe comments that universities felt difficulty in what and how much evidences should be prepared.

Table5: Summary of descriptive statistics for Questions 3 between 2 cycles (N=52)

Questions	First Cycle		Second Cycle	
	M	*SD*	*M*	*SD*
(1) The workload required for the evaluation				
1. Preparing the self-evaluation form	4.45	.72	4.46	.57
2. Addressing the "Checkpoints During Site Visits" presented before site visits **	3.41	.66	3.90	.71

3.	Advance preparations for site visits **	3.39	.63	3.77	.64
4.	The day of site visit *	3.18	.55	3.46	,66
5.	Statement of objection	2.82	.71	2.94	.60
(2)	Given the workload, was the timeframe for evaluation set by NIAD-QE appropriate?				
1.	Completing the "Checkpoints During Site Visit" presented before site visits	2.73	.66	2.50	.80
2.	Advance preparations for site visits	2.76	.70	2.71	.69
3.	The day of site visit	2.98	.31	2.88	.62
4.	Statement of objection	2.82	.62	3.11	.43
(3)	The amount of effort required for the evaluation was appropriate for the evaluation.				
1.	The amount of effort required for the evaluation was appropriate for quality control of your institution's educational and research activities.	3.71	.72	3.81	.56
2.	The amount of effort required for the evaluation was consistent with the progress of improvement of your institution's educational and research activities.	3.82	.71	3.77	.54
3.	The amount of effort required for the evaluation was consistent with your institution's sup-port for public understanding and educational and research activities.	3.65	.71	3.31	.77
(4)	The evaluation schedule				
1.	The timing of the submission of the self-evaluation form (before the end of June) was suitable [a)]	1.69	.46	1.80	.40
2.	The timing of site visits for the survey (early October to mid-December) was suitable. [a)]	1.96	.19	1.98	.14

[a)] *two-points scale*
*Notes. * $p < .05$, ** $p < .01$*

3.4 Orientation meetings and training sessions

This section includes questions about the orientation meetings and training sessions for self-evaluation. Only Question 4-9 showed a significant difference in each cycle ($t (51) = 2.26$, $p < .05$). In Question 4-9, positive answers are reduced in the second cycle; however, the overall ratings are still positive in each cycle. Other questions also showed positive answers in each cycle and did not show significant differences: (4-1, $t (51) = .80$, $p = .43$; 4-2, $t (51) = .93$, $p = .36$; 4-3, $t (51) = .64$, $p = .53$; 4-4, $t (51) = 1.07$, $p = .29$; 4-5, $t (51) = .88$, $p = .38$; 4-6, $t (51) = 1.05$, $p = .30$; 4-7, $t (51) = -.78$, $p = .44$; 4-8, $t (18) = -.77$, $p = .45$).

The description of the open-ended questions showed that the universities requested more detailed explanations for the standards and viewpoints that presented difficulty during the self-evaluation process.

Table6: Summary of descriptive statistics for Questions 4 between 2 cycles (*N*=52)

Questions	First Cycle		Second Cycle	
	M	*SD*	*M*	*SD*
1. The materials distributed during orientation were easy to understand.	3.87	.50	3.84	.44
2. The contents of orientations were easy to understand	3.87	.50	3.82	.47
3. The contents of orientations were useful.	4.02	.71	4.00	.56
4. The materials distributed during training for self-evaluating supervisors were easy to understand.	3.85	.63	3.78	,62
5. The training contents for self-evaluating supervisors were easy to understand.	3.83	.51	3.78	.60
6. The training contents for self-evaluating supervisors were useful.	4.04	.50	3.96	.48
7. The booklet of implementation points for self-evaluation provided by NIAD-QE was useful.	4.08	.49	4.25	.39
8. The visit for orientation made by NIAD-QE was useful. [a]	4.11	.72	4.26	.71
9. The actions taken by NIAD-QE's administrative supervisor at orientation meetings and training sessions (e.g., responses to questions) were appropriate. *	4.25	.58	4.02	.61

[a] *N*=19

Notes. * *p* < .05

3.5 Evaluation results (evaluation report form)

Subsection 5-(1), the questions about universities' satisfaction with the evaluation results. In this subsection, all answers were positive through two cycles in general. In Question 5-(1)-3 (t (51) = 2.27, p < .05) and 5-(1)-9 (t (51) = 2.26, p < .05), positive answers are significantly reduced in the second cycle. However, the overall ratings are still positive in each cycle.

Other questions in this subsection also showed positive answers in each cycle and did not show a significant difference: (5-(1)-1, t (51) = 1.43, p = .16; 5-(1)-2, t (51) = 1.00, p = .32; 5-(1)-4, t (51) = 1.97, p = .06; 5-(1)-5, t (51) = 1.83, p = .07; 5-(1)-6, t (51) = .30, p = .77; 5-(1)-7, t (51) = 0.00, p = 1.00; 5-(1)-8, t (51) = .73, p = .47).

Subsection 5-(2) includes two questions about the situation of universities' public disclosure of the self-evaluation reports and evaluation reports. Neither of the two questions showed a significant difference (5-(2)-1, χ^2 (1) = 1.01, p = .31; 5-(2)-2, χ^2 (1) = 0.00, p = 1.00). The answer clarified that almost all the universities disclose these documents.

Question 5-(3)-1 concerns the appropriateness of publicizing the evaluation results through the mass media. There was no significant difference between two cycles (t (38) = .59, p = .55). However, the overall ratings were negative in each cycle. Although there was no significant difference, negative descriptions in the open-ended questions field have increased. The typical descriptions in the first cycle were about specific reports by the mass media (e.g., "Mass media

doesn't have enough understanding about the CEA to report" and "The tendency of the attention to the negative point in the report of the mass media makes it difficult to get social support for accredited universities."). Several comments to the NIAD-QE, requiring promotion of activities to cultivate the significance of the CEA and the social importance of encouraging interest and cooperating with MEXT, are reported. In the second cycle, the typical descriptions are also negative; the following two answers are representative: "In the first place, the report from the mass media itself was insufficient," and "After the disclosure of the results of CEA, there were no special comments from the mass media."

Table7: Summary of descriptive statistics for Questions 5 between 2 cycles (*N*=52)

	First Cycle		Second Cycle	
Questions	*M*	*SD*	*M*	*SD*
(1) The contents of the valuation report form				
1. The contents of the evaluation report form were appropriate for the quality control of your institution's educational and research activ-ities.	4.12	.51	4.00	.48
2. The contents of the evaluation report form were appropriate for encouraging improve-ments to your institution's educational and research activities.	4.12	.61	4.00	.59
3. The contents of the evaluation report form were appropriate for gaining public under-standing and support for your institution's edu-cational and research activities. *	3.85	.72	3.56	.60
4. The contents of the evaluation report form were appropriate for your institution's objectives.	4.15	.53	3.94	,53
5. The contents of the evaluation report form were in line with your institution's current situation.	4.23	.50	4.08	.47
6. The contents of the evaluation report form considered your institution's rules, resources, and systems.	3.88	.72	3.85	.69
7. The contents of the evaluation report form offered fresh perspectives on educational and research activities.	3.60	.60	3.60	.63
8. The contents and composition of the evalua-tion report form were easy to understand.	4.08	.58	4.00	.59
9. The overall contents of NIAD-QE's evalua-tion report form were appropriate. *	4.29	.45	4.04	.39
(2) The publication of the self-evaluation and evaluation report forms				
1. Your institution has published the completed self-evaluation form for this evaluation online. [a), b)]	2.00	.00	1.97	.16
2. Your institution has published the completed evaluation report form online. [a), b)]	1.97	.16	1.97	.16

(3) Release of information on evaluation results to the mass media

| 1. | Your institution has received appropriate coverage of the evaluation results by the mass media. [a] | 3.10 | .93 | 3.00 | .55 |

[a] *N=39,* [b] *two-points scale*

*Notes. * p < .05*

3.6 The effects and impacts from evaluation

Section 6 includes questions about the effects and impacts from CEA. Subsection 6-(1) investigates the effects and outcomes resulting from the self-evaluation process. Subsection 6-(2) includes questions about how universities perceive the effects and outcomes of the results of CEA. In this section, only Questions 6-(2)-7 showed significant differences between two cycles (t (51) = 2.10, $p < .05$), whereas there were no significant differences in the other questions (6-(1)-1, t (51) = 1.26, $p = .21$; 6-(1)-2, t (51) = .75, $p = .46$; 6-(1)-3, t (51) = .74, $p = .46$; 6-(1)-4, t (51) = 1.42, $p = .16$; 6-(1)-5, t (51) = -.73, $p = .47$; 6-(1)-6, t (38) = .00, $p = 1.00$; 6-(1)-7, t (51) = .00, $p = 1.00$; 6-(1)-8, t (51) = .74, $p = .46$; 6-(1)-9, t (51) = .16, $p = .88$; 6-(1)-10, t (38) = .52, $p = .61$; 6-(2)-1, t (51) = .35, $p = .73$; 6-(2)-2, t (51) = .75, $p = .46$; 6-(2)-3, t (51) = 1.38, $p = .17$; 6-(2)-4, t (51) = .98, $p = .33$; 6-(2)-5, t (51) = .00, $p = 1.00$; 6-(2)-6, t (38) = 1.42, $p = .16$; 6-(2)-8, t (51) = .97, $p = .34$; 6-(2)-9, t (51) = 1.27, $p = .21$; 6-(2)-10, t (51) = 1.59, $p = .12$; 6-(2)-11, t (38) = 1.76, $p = .09$; 6-(2)-12, t (38) = -.31, $p = .76$; 6-(2)-13, t (51) = 1.90, $p = .06$; 6-(2)-14, t (51) = 1.13, $p = .26$; 6-(2)-15, t (51) = .82, $p = .42$).

In Question 6-(2)-7, asking if the result from the evaluation facilitated the improvement of management, positive answers are seen in the first cycle. This result was significantly reduced in the second cycle and shifted to neutral answers.

Although there were no significant differences, the overall trend in this section revealed that the universities' perception toward the effectiveness of evaluation results became negative in the second cycle.

To sum up the answers to the questions in this section, we can see two findings.
First, as can be seen from Questions 6-(1)-3, 6-(1)-4, 6-(1)-9, 6-(1)-10, 6-(2)-3, 6-(2)-4, and 6-(2)-10, the effects and outcomes of the evaluations on faculty members and staff resulting from the CEA have decreased in the second cycle. The free description that represents this comment is about the narrow effects of the CAE on the faculty members and staff. The faculty members and staff who were concerned with the evaluation process can recognize the importance of the CEA, whereas most people do not know about it at all.

Secondly, in Questions 6-(2)-1, 6-(2)-2, 6-(2)-5, and 6-(2)-12, positive answers are seen in each cycle.

Table8: Summary of descriptive statistics for Questions 6 between 2 cycles (*N=52*)

Questions	First Cycle		Second Cycle	
	M	*SD*	*M*	*SD*
(1) The effects and impacts from self-evaluation				
1. An overall understanding of your institution's educational and research activities could be achieved.	4.31	.64	4.17	.47

2.	Understanding of future issues regarding your institution's educational and research activities could be achieved.	4.13	.48	4.06	.46
3.	Faculty and staff understood the importance of systematic management of educational and research activities	3.38	.71	3.29	.60
4.	All faculty members gained an enhanced awareness of engagement in educational and research activities.	3.38	.65	3.21	,60
5.	Improvements to your institution's educational and research activities were encouraged.	3.79	.66	3.87	.39
6.	The evaluation was useful for formulating your institution's plans for the future. [a)]	3.41	.54	3.41	.54
7.	Improvements to your institution's management were encouraged.	3.52	.64	3.52	.64
8.	Your institution's particular practices were promoted	3.42	.69	3.33	.61
9.	Faculty and staff gained an understanding of the importance of self-evaluation.	3.42	.69	3.40	.66
10.	Knowledge and skills of faculty and staff in evaluation improved.	3.51	.59	3.44	.63

(2) The effects and impacts from NIAD-QE's evaluation results

1.	An overall understanding of your institution's educational and research activities could be attained.	4.13	.59	4.10	.49
2.	An understanding of future issues concerning your institution's educational and research activ-ities could be attained.	4.12	.51	4.04	.55
3.	Faculty and staff gained an understanding of the importance of systematic management of educational and research activities.	3.63	.65	3.48	.60
4.	All faculty members attained an enhanced awareness of engaging with issues in educa-tional and research activities.	3.58	.63	3.44	.66
5.	Improvements to your institution's educational and research activities were encouraged.	3.88	.54	3.88	.54
6.	The evaluation was useful in formulating your institution's plans for the future. [a)]	3.54	.55	3.36	.58
7.	Improvements to your institution's management were encouraged. *	3.75	.58	3.50	.64
8.	Your institution's particular practices were promoted.	3.58	.63	3.46	.60
9.	Faculty and staff attained an understanding of the importance of self-evaluation.	3.54	.72	3.38	.65

10. Faculty and staff knew and understood the contents of the evaluation results.	3.67	.54	3.50	.54
11. Knowledge and skills of faculty and staff in evaluation approaches and methodology improved. [a]	3.69	.69	3.41	.71
12. The quality of your institution's educational and research activities has been assured.	3.90	.66	3.94	.50
13. Your institution has obtained student's support and understanding.	3.38	.62	3.13	.68
14. Your institution has obtained widespread public support and understanding.	3.50	.67	3.35	.70
15. Your institution has drawn on other universities' good practices based on their evaluation results.	3.73	.62	3.63	.62

[a] *N*=39

Notes. * $p < .05$

3.7 The effects and impacts from your last certified evaluation

Section 7 includes questions about a longitudinal study and the impact from the first cycle CEA conducted approximately seven years ago. We analyzed whether there were differences between responses to the question asking about the effectiveness of the new CEA (Question 6-2) and the previous CEA (Question 9). All three questions 9-1, 9-2, and 9-3 showed significant differences compared with the corresponding Questions 6-(2)-12, 6-(2)-5, 6-(2)-14, respectively (9-1, t (48) = -5.96, $p < .01$; 9-2, t (48) = -3.50, $p < .01$; 9-3, t (48) = -2.92, $p < .01$). When universities rated 4 or 5 out of 5 scales in Question 9, the survey required them to provide detailed descriptions on what the beneficial outcomes and impacts were. In representative comments about 9-2, universities reported the perception that the CEA encouraged the improvement of their educational and research activities because they applied feedback on their improvement in quality.

Table9: Summary of descriptive statistics for Questions 9 in the second cycle (*N*=49)

Questions	*M*	*SD*
1. The previous certified evaluation produced beneficial effects from and impacts of quality control of your institution's educational and research activities.	3.37	.49
2. The previous certified evaluation produced beneficial effects from and impacts of the support for improvements to your institution's educational and research activities.	3.56	.50
3. The previous certified evaluation produced beneficial effects and impacts of public understanding and support for your institution's educational and research activities.	2.94	.35

3.8 NIAD-QE's certified evaluation process in comparison with previous evaluations

Section 8 includes questions about whether NIAD-QE's CEA process has improved from the first CEA cycle conducted approximately seven years ago.

Table10: Summary of descriptive statistics for Questions 10 in the second cycle

Questions	N	M	SD
1. The composition and content of evaluation standards and perspectives have become more appropriate for certified certification evaluation.	51	3.76	.55
2. Because of the evaluation standards and perspectives, it has become possible to prepare a more appropriate self-evaluation form.	51	3.67	.55
3. The practical contents and conduct of visits for the survey have become more appropriate.	51	3.45	.58
4. The workload required for the evaluation and period of operations set by NIAD-QE have become more appropriate.	50	3.18	.52
5. The amount of effort required for the evaluation is now more consistent with the purpose of the certified evaluation.	50	3.32	.47
6. Orientation meetings and training sessions have become easier to understand and more useful.	49	3.53	.58
7. The contents of the evaluation report form are now more consistent with the purpose of the certified evaluation.	51	3.53	.61
8. Your institution is now more positive about publicizing its self-evaluation and evaluation report forms.	49	3.41	.57
9. Media coverage of the evaluation results has become more appropriate.	46	3.00	.30
10. You have had greater effects and impacts from self-evaluation.	50	3.32	.47
11. You have had greater effects and impacts from NIAD-QE's evaluation.	50	3.34	.52

4 Discussion

The study's purpose was to explore what accredited universities perceived about the effectiveness, impact, and challenges of the CEA system and practices and how these have been changed between the first and the second cycles. In order to examine these issues, the survey results were analyzed by two approaches. First, the data was compared between the first and the second cycles. Second, we analyzed whether there were differences between responses to a question asking about the effectiveness of the new CEA (Question 6-(2)) and a question about the effectiveness of the previous CEA (Question 9). Results from statistical analyses between the two cycles

showed that responses were different for all questions, but have changed significantly for nine out of 87 questions. Although universities' perceptions of the CEA have not changed sharply in general, the differences in these nine questions reveal important facts, which can lead to the growth of CEA system. In contrast, there was a significant difference between the response to the question asking about the effectiveness of the new CEA (Question 6-(2)) and improvement from the previous one (Question 9). We will analyze and discuss the results following three themes: "effectiveness in helping quality enhancement," "gaining public understanding and support," and "evaluation exhaustion (workload)."

Effectiveness in helping quality enhancement:

As previously mentioned, the survey in the second cycle asked universities in two ways about CEA effectiveness for quality improvement; one question (6-(2)-5) referred to the immediate impact of CEA in the second cycle and another Question (9-2) asked about a longitudinal impact from the first cycle CEA conducted approximately seven years ago (Question 9-2). The results showed that the majority of universities had a relatively positive perception of the effectiveness of the CEA for improvement in university quality. Specifically, when universities rated 4 or 5 on a 5-point scale in Question (9-2), the survey required them to provide detailed descriptions of the beneficial outcomes and impacts. In their comments, universities reported perceptions that the CEA encouraged the improvement of their educational and research activities because it applied feedback on their improvement in quality. This perception is consistent with the data that approximately 70% of all 53 universities applied "improvement needed" in the accreditation results to their quality improvement in the second cycle, in response to the first cycle evaluation result [15].

Regarding Question 6-(2)-5, asking whether NIAD-QE's evaluation results have helped im-prove universities' educational and research activities, the responses were positive in both cycles and not significant, suggesting that universities perceived the CEA as an effective way to im-prove their educational and research activities. To Question (9-2), which appeared only in the survey for the second cycle, asking whether receiving the CEA approximately seven years ear-lier had helped universities to produce beneficial outcomes or impacts for improvements of their educational and research activities, their response was more than average (3.56)—but it was significantly lower than the responses in Question (6-(2)-5), which asked about the immediate impact of CEA. It is conceivable that a positive perception was obtained because the survey was conducted immediately after the CEA was performed; universities had just experienced pre-paring for their own self-evaluation reports and receiving the accreditation feedback.

It is also important to note that "improvement needed" in the first cycle primarily pointed out student admission (Standard 4, especially the management of graduate school student numbers), facilities (Standard 8), and academic staff and educational supporting staff (Standard 3); few pointed out the areas of credit substantiation or an internal quality assurance system which would require the development or improvement of a longer period of time and a heavier workload. In the second cycle, many "improvement needed" points referred to matters which would directly lead to improvement in the quality of universities' educational and research activities, such as credit substantiation, and establishment of an internal quality assurance system [15]. As self-initiative is required to apply feedback to universities in such domains, this would be a major challenge for them.

It is also worth noting that the awareness of faculty members and staff with regard to the im-portance of the CEA self-evaluation process (Questions 6-(1)-3, 6-(1)-4, 6-(1)-9, 6-(1)-10), the effects and outcomes of receiving the CEA results (6-(2)-3, 6-(2)-4, and 6-(2)-10) tended to decline in the second cycle, though these declines were not significant. Related comments demonstrated that only a part of faculty members and staff who had actually been involved in the

evaluation process could recognize the importance of the CEA, whereas those who were not concerned of any quality assurance procedures were not aware of it at all. There have been concerns that the CEA practice is perceived as a technical matter that requires the involvement of only a limited number of members, although the newly revised CEA system emphasizes internal quality assurance, expecting more involvement of a wide range of constituents at universities, programs, or courses.

Gaining public understanding and support:

Gaining public understanding and support for universities significantly decreased in the second cycle (Question 5-(3)-1). As reviewed earlier, in the first cycle overview, gaining public understanding and support, one of the NIAD-QE objectives, still requires more initiative and effort. The evaluation standards and viewpoints were revised, placing more emphasis on "public information on teaching and learning"; however, improvement was not observed in the second cycle. Frequent comments provided by universities in the first cycle referred to how mass media had treated the CEA results in the public, such as "the mass media does not fully understand nor interpret the CEA reports," and "mass media overemphasized a negative result in the accreditation report, which would make it difficult to obtain the social understanding and trust for accredited universities. In the second cycle, the typical comments in universities' descriptions were still negative, such as "the mass media never reacted to the CEA even though we released the accreditation results to the public." These comments suggested that NIAD-QE and universities need to devise a mass media strategy in order to make appropriate information available to the public and to receive social understanding and support. Further discussion about this issue is needed among NIAD-QE and universities. For example, sharing universities' good practices with the media, or developing user-friendly reports so that the public feels more comfortable about understanding the CEA, could be some of the solutions.

Evaluation exhaustion:

The workloads for site visits significantly increased in the second cycle (3-(1)-2, 3-(1)-3, 3-(1)-4). It is likely that this tendency reflected the government and quality assurance institutions' recent emphasis on submitting concrete and detailed evidence of quality assurance. Universities had to collect a wide variety of data or resources from each department, program, or course within three weeks. This means that the larger a university is, the larger its workloads become. To avoid this problem, more effective development of institutional management systems is encouraged, leading to effective governance at each unit (e.g., department, program, and course), including data assessing student learning outcomes, summaries of course evaluations, or portfolios on a regular basis. Related to this issue, in Question 6-(2)-7, which asked whether the evaluation result helped the university improve institutional management, positive responses were observed in the first cycle; however, these were significantly reduced in the second cycle and shifted to neutral reaction. The data to be evidence of the challenge that management handled well.

From descriptive comments, this study showed universities felt difficulty concerning what and how much evidences they should prepare. Developing databases for collecting quantitative data of higher education could be one solution for this problem. Setting the standardized key performance indicators and evidence in each standard and viewpoint would make it possible to reduce workloads. It should be understood that adequate communication between universities and NIAD-QE is needed in this process. As we noted at the beginning, although much research has provided indicators and evidence for student learning outcomes by analyzing university

evaluation reports [7][8][9], the indicators and evidence for other standards and viewpoints has had little investigation. Therefore, it is expected the further research will reduce workloads.

Other, needs for improvement in terms of the orientation meeting and training sessions:

The response regarding how helpful orientation meetings and training sessions (e.g., responses to questions) offered by NIAD-QE were to universities was lower in the second cycle than that of in the first cycle, although still positive (4.0). Due to the lack of university comments, it is difficult to explain why the change was found. Our study found that universities request more detailed explanations for the evaluation standards and viewpoints for which they encountered difficulties in preparing data and evidence. It is necessary for quality assurance institutions to provide useful evidence and resources through orientation meeting and training sessions for CEA. It is critical that universities and quality assurance institution share information to assure quality.

We acknowledge that there are several limitations to our study.

First, the study aimed at revealing the perception of universities of the CEA, based on the questionnaire survey. The questionnaire method is one of useful approach for measuring indirect outcomes through the universities' perception. However, it does not cover all the outcomes, including the results from the direct changes of the universities induced by the CEA. The results of this study do not reflect the actual effect of CEA. Further research is needed using objective data that reflects the direct changes in the universities that do result from the CEA.

Second, in this study, subject universities were mainly national universities. Therefore, the data included the trait of national universities. The advantage to confining our discussion to the national universities is that the data is controlled, without confounding variables influenced by differences in university characteristics. Meta-analysis of the data of the questionnaire survey in other accreditation institutions (JUAA and JIHEE) makes it possible to generalize the result for all the universities in Japan.

Third, it should be noted that there is a possibility of artifacts induced by the difference of the rater from the first cycle and second cycle. Especially in the second cycle survey, many raters may not have experienced the previous CEA. However, it can be assumed that the culture and history of a university is universal; we regarded them as the same sample in spite of the differences between the raters in the two cycles.

Although there are limitations, this study contributes to making a better CEA system. We provided useful findings by analyzing changes in the universities' perception of the CEA through the annual questionnaire survey. It is hoped that the findings that have been presented in this paper will contribute to improvement in both internal and external quality assurance systems.

References

[1] JUAA, "'Daigaku Hyoka (Ninshohyoka)' no Yukosei ni Kansuru Chosa," Japan University Accreditation Association, 2012;
http://www.juaa.or.jp/images/publication/pdf/other/daigakuhyouka_report.pdf/ (accessed 2016-11-20)

[2] NIAD-UE, "Evolving Institutional Certified Evaluation and Accreditation," National Institution for Academic Degrees and University Evaluation. 2013;
http://www.niad.ac.jp/n_hyouka/jouhou/__icsFiles/afieldfile/2013/05/22/no6_12_soukatsu.d

aigaku.pdf/ (accessed 2016-11-20)

[3] JIHEE, "Ninshohyoka ni Kansuru Chosakenkyu," Japan Institution for Higher Education Evaluation, 2009; http://www.jihee.or.jp/publication/pdf/research_report/h20_monka_jigyou.pdf / (accessed 2016-11-20)

[4] JUAA, "'Daigaku Kijun' oyobi Sono Kaisetsu," Japan University Accreditation Association, 2010; http://www.juaa.or.jp/updata/news/411/20160219_156663.pdf / (accessed 2016-11-20)

[5] NIAD-UE, "Institutional Certified Evaluation and Accreditation of Universities Standards for Evaluation and Accreditation of Universities: 2012-2019," National Institution for Academic Degrees and University Evaluation. 2011; http://www.niad.ac.jp/n_hyouka/daigaku/__icsFiles/afieldfile/2016/05/24/no6_1_1_daigaku2kijun29.pdf / (accessed 2016-11-20)

[6] JIHEE, "Daigaku Hyoka Kijun," Japan Institution for Higher Education Evaluation, 2015; http://www.jihee.or.jp/achievement/college/pdf/hyokakijyun140902.pdf / (accessed 2016-11-20)

[7] A. Noda, and S. Shibui, "Substantialization of the Credit Hour System and Certified Evaluation and Accreditation," Research on Academic Degrees and University Evaluation, no. 17, 2016, pp. 19-33.

[8] A. Noda, and S. Shibui, "The Challenges for Accreditation to Enhance Credit Hour System in Japan," Proc. The 11th Hawaii International Conference on Education, 2013, pp.921-945.

[9] S. Shibui, N. Takahashi, and A. Noda, "Visualization of the Cognitive Dimensions for Evaluating Universities By Means of the Content Analysis of Evaluation Reports," Abstract. International Meeting of the Psychonomic Society, 2016, pp.192-193.

[10] T. Saito, and T. Hayashi, "For Further Improvement of Universities: Evaluation of the Evaluation Program Implemented by National Institution for Academic Degrees and University Evaluation in Japan," Japanese Journal of Evaluation Studies, vol. 7, no. 1, 2007, pp. 33-46.

[11] K. Ogiue, "System of Certified University Evaluations: Its Problems and Its Future," University Evaluation Review, no. 8, 2009, pp. 43-51.

[12] I. Amano, "Ninshohyoka no Gendankai," IDE: Gendai no kotou kyoiku, no. 504, pp.4-11.

[13] H. Sato, M. Mori, E. Takata, T. Kominato, and M. Sekiguchi, "Challenges that Institutional Researchers Face in the Scene of University Evaluation: A Qualitative Analysis of Questionnaire," Research on Academic Degrees and University Evaluation, no. 9, 2009, pp. 63-77.

[14] NIAD-UE, "Overview of the Quality Assurance System in Higher Education Japan, " National Institution for Academic Degrees and University Evaluation. 2009; http://www.niad.ac.jp/n_kokusai/info/japan/overview_jp_j.pdf/ (accessed 2016-11-20)

[15] N. Takahashi, S. Shibui, and A. Noda, "An Analysis of Japanese Universities' Quality Im-provement Initiatives Based on Accreditation," Abstract. The 15th Hawaii International Conference on Education, 2017, pp. 148.

[16] S. Kim, T. Hayashi, and T. Saito, "Model of Institutional Quality Enhancement Promoted by "Certified Evaluation and Accreditation": Comparative Analysis among Different Types of Higher Education Institutions," Research on Academic Degrees and University Evaluation, no. 9, 2009, pp. 19-42.