

Outcomes assessment and its role in self-reviews of undergraduate education: in the context of Japanese higher education reforms since the 1990s

Takeshi Kushimoto

Published online: 6 August 2009
© Springer Science+Business Media B.V. 2009

Abstract Under the recent state of higher education, “Outcomes Assessment” has become a familiar term in Japan as in other nations all over the world. However, actual conditions of outcomes assessment and its contribution toward educational improvement are not always obvious. Thus, this article attempts to clarify: (1) Japanese higher education reforms since the 1990s focusing on assessment and evaluation; (2) the influence of reforms on outcomes assessment implemented by higher education institutions; (3) how much the results of outcomes assessment are used in self-reviews of undergraduate education; and (4) whether application of outcomes assessment contribute to educational improvement. The results of national survey conducted in Japan reveal that assessments are clearly being carried out, but may not have been supporting the improvement of education.

Keywords Learning outcome · Outcomes assessment · Self-review · Japanese higher education · Logistic regression analysis

Background and purpose

Worldwide trend in outcomes assessment

Under the recent state of higher education, “Outcomes Assessment” has become a familiar term not only in the United States but also in other nations around the world. Several factors may be causing this trend, but mainly two reasons stand out.

The first one is a call for accomplishing accountability. Since higher education institutions (HEIs) are run by public money, they have the responsibility to show that the money spent is producing benefits. Especially under times of depression, these types of demands intensify. Outcomes assessment could be a tool used to answer to such demands, by tendering information about student learning (Dill 2000).

T. Kushimoto (✉)
Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan
e-mail: ktakeshi@tmu.ac.jp

The second reason relates to the so-called “paradigm shift in higher education” (Barr and Tagg 1995). When the focus of education moves from “instructions by the teacher” to “learning by the student”, the necessity to understand student learning through outcomes assessment is naturally increased. Assessment results therefore become a basis for pedagogy and curriculum refinement (Huba and Freed 2000). It is at this stage where outcomes assessment becomes an essential function for improving undergraduate education.

The two-way interpretation of assessment needs such as this has been used for more than a decade, as one can see in the “external and internal demands” dichotomy in Ewell (1997) and others. However, it is still valid, and also useful in explaining the case for Japan. Briefly put, it is reasonably predictable that outcomes assessment has been introduced due to external demands, but there remains uncertainty as to whether the assessments are being utilized in self-reviews of undergraduate education and contributing to the improvement of undergraduate education.

Research questions and organization of this article

The difficulty in connecting assessment (grasping information) and evaluation (using that information in the process of decision making judgment) is recognized even in the U.S., which has a much longer history of outcomes assessment than Japan (López 2004). Investigating the situation in Japan may therefore be of little value since outcomes assessment is a fairly new phenomenon there. However, it is meaningful to demonstrate, with objective data, how policy reforms affect the behavior of HEIs, in order to record the facts and offer a reference point for comparative studies.

Thus, this article will attempt to clarify: (1) Japanese higher education reforms since the 1990s focusing on assessment and evaluation, (2) the influence of reforms on outcomes assessment implemented by HEIs, (3) how much the results of outcomes assessment are used in self-reviews of undergraduate education, and (4) whether application of outcomes assessment contribute to educational improvement.

The contents of this article are organized into three sections. The first section depicts how Japanese higher education has been reformed since the 1990s. Among various attempts, two events are selected according to their concern for assessment and evaluation. In the subsequent section, data from a national survey demonstrates the degree of implementation of outcomes assessment in Japan. The survey also reveals how far the adoption of assessment results in self-review has spread, and how much impact the assessment has had on educational improvement. The final section discusses the results of the survey, and also identifies keys for further research through theoretical interpretations of the discussed points.

Higher education reforms in Japan

Overview of Japanese higher education system

Japanese higher education system, which started in 1870s, has consistently expanded its scale especially after the WW II. There are 765 universities and nearly 3 million students are enrolled in them by the statistics of 2008. Three characteristics of Japanese higher education are presented here.

First, private sector is fairly large. Seventy seven percent of 765 universities are private and the ratio of students going to private universities reaches 73%. These percentages teach us that the expansion of higher education has been supported by private sector in Japan.

Second, Japanese universities have been said, “hard to enter, easy to graduate”. Since the number and ratio of high school students advancing to higher education had been increased until 1990s, the screening function of entrance examinations had been emphasized. On the other hand, people tended not to make a point of learning outcomes in university education because the lifetime employment and OJT were premised by Japanese companies.¹ These situations let people conceive that universities are hard to enter; but once entered, one can enjoy a time of freedom till graduation.

Third, partly as a consequence of the second characteristic written above, many reforms have been taking place since 1990s. The underlying idea of the reforms is “deregulation and self-responsibility”, which remind us the term “Evaluative State” explained by Neave (1988). Incorporation of national universities and the evolution of evaluation system are major example of the reforms. In the following parts, details of the latter will be illustrated.

Revision of law in 1991

The revision of higher education law (Daigaku-Secchi-Kijyun) in 1991 is seen as the first biggest reform that followed the one after WW II. With this revision, universities came to have greater discretion over curriculum planning as a result of the abolishment of minimum requirements on general education (Amano and Poole 2005). Moreover, the title of “bachelor” was officially recognized as a “degree”, and a regulation concerning the types of bachelor degrees was removed to emphasize a commonality in undergraduate education.

In addition to these alterations, the revision also provided a new regulation relating to assessment and evaluation. This new regulation put into place an unforced obligation to implement self-review of university activities. Instead of loosening the regulation on curriculums, requirements were placed on universities to constantly check their activities and enhance the quality of education by themselves (Kitamura 1997).

In 1999, the word “unforced” was removed from the provision, and furthermore, universities were required to make the results of self-review public from then on. Following such progress, 98% of 731 universities have conducted self-reviews during the period from 1999 to 2006 (Ministry of Education 2008). This figure suggests that self-reviews have become a kind of routine undertaking for all Japanese universities.

Nevertheless, outcomes assessment has not always been used in self-reviews until recently as universities have been reluctant to take untraditional actions. The next part will explain how trial evaluations by NIAD (National Institution for Academic Degrees) prompted a change in awareness and behavior.

Reorganization of NIAD in 2000

NIAD was founded in 1991. The original role of NIAD was to award degrees to those who completed undergraduate education in unusual ways. The function of university evaluation was added in 2000 according to the report by the University Council (1998) which proposed the importance of having a third-party organization for university evaluation (Yonezawa 2002).²

¹ In fact, about 80% students graduate university in 4 years—standard enrolment period of most majors other than ones related to medical activities.

² Although NIAD changed its English name to NIAD-UE (University Evaluation), the abbreviation “NIAD” is used throughout this article. For more information about NIAD, see http://www.niad.ac.jp/index_e.html.

Table 1 Results of trial evaluations

Levels	Organization	Curriculum	Pedagogy	Outcomes
Suitable to objectives	0	1	0	0
Almost suitable to objectives	43	78	20	4
Needs some improvement	49	16	72	63
Needs much improvement	3	0	3	28
Should be reconsidered	0	0	0	0
Tentative mean score	2.42	2.84	2.18	1.75

Note: Numerical values represent the number of universities. Tentative mean score is calculated as 4 for “Suitable to objectives” and 0 for “Should be reconsidered”. It should be noted that NIAD explained that the mean score does not make sense as levels were given relative to each institution’s objectives

From 2000 to 2004, NIAD carried out trial evaluations mainly of national universities. Although there were various types of evaluation in terms of content, i.e., education or research, and the targeted organization, i.e., the whole institution or individual faculty, the evaluation of liberal education (Kouyou-Kyouiku) received the most attention.³ The results of this evaluation, which was carried out in four areas, are shown in Table 1.

The reason for such attention, simply stated, was the unfavorable result of “outcomes”. Yet, although this seems to suggest that liberal education does not work well, research on the process of this evaluation asserts that such an interpretation is not accurate. Kushimoto (2004) indicates that the results of the evaluation revealed “the absence of information about outcomes of liberal education, rather than insufficient effect of it”. In other words, it could be said that the trial evaluations by NIAD made people aware of the necessity of outcomes assessment.

After this event, outcomes assessment began to attract attentions explicitly in the discussions and practices of evaluations in Japan.⁴ Nonetheless, it is obvious from the progress shown above that interests in outcomes assessment come from the aspect of accountability rather than improvement. This gives us to predict the difficulty in connecting outcomes assessment to self-review which aims to educational improvement.

Outcomes assessment and its role

Data

The data analyzed in this section was collected through “the survey on the evaluation and improvement of education”. The survey was conducted at the beginning of FY 2006 with 1,871 faculties which had undergraduate courses. 580 faculties replied. Thus the response rate was approximately 31%.

³ The definition of liberal education is controversial in Japan as is in other countries. In the context of this article, it is enough to be interpreted as education whose objectives are valid regardless of students’ major or subject of study.

⁴ From 2004, mandatory system of third-party evaluation (Ninsyo-Hyouka-Seido) has started in Japan. Under this new system, all HEIs have to receive an evaluation by a nationally certified organization once every 7 years. NIAD and other certified organizations set outcomes assessment as one of the evaluation standards.

The questionnaire was constructed from five parts: traits of faculty, framework of self-review, methods of self-review, effects of self-review and free opinions of the respondent about self-review. Questions were determined with consideration of preceding research such as that by the Research Institute for Higher Education Hiroshima University (1999), Kushimoto (2006) and Peterson et al. (1999).

Implementation of outcomes assessment

The section on “methods of self-review” in the questionnaire asks what kinds of outcomes assessment are being carried out and from when. The answers are displayed in Fig. 1. The three periods indicating the start of assessment correspond to those events mentioned in the previous section.

Assessment methods can be divided into four types in accordance with two different perspectives. On the one hand, one can see the difference between direct assessments and indirect ones (Palomba and Banta 2001). While direct assessment somehow tries to measure outcomes themselves, indirect assessment utilizes subjective opinions and data that can reflect something other than learning outcome (e. g., retention and graduation rates). On the other hand, assessments are categorized as either comparable or incomparable. Owing to the commonality of indicators, the results of comparable assessment can be used to compare across institutions for which the assessments are done.

Grasping the destinations of graduates is one method of indirect-comparative (I-C) assessments. The destinations do not always reflect student learning outcomes during undergraduate education since they may be determined by the name of the university and so on. The survey revealed that many faculties of Japanese universities had collected the data before 1990.

The next three methods in the figure are direct-comparative (D-C) assessments. Since the past, universities have been familiar with national examinations, which are necessary to obtain licenses for teaching and medical professions and usually taken near the end of undergraduate courses. On the contrary, tests which are unrelated to the students’ major

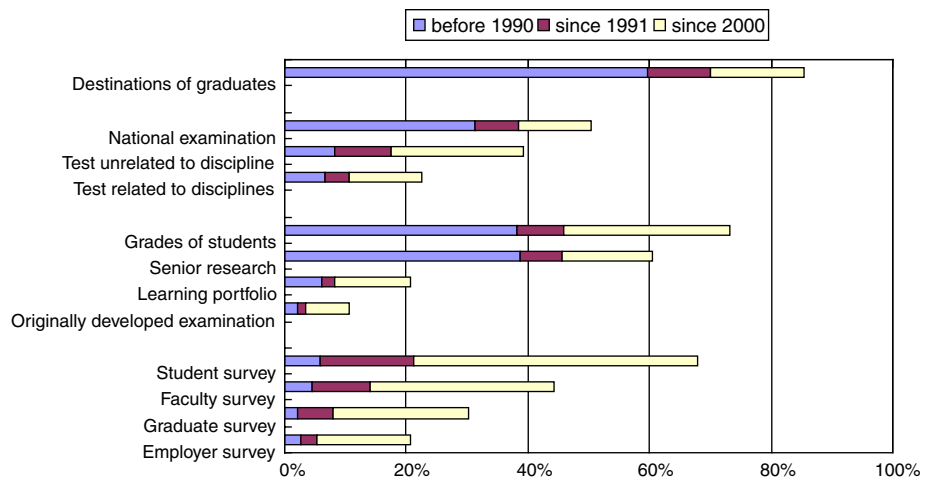


Fig. 1 Outcomes assessment prevalence

(e.g., TOEIC and qualifications for general skills) and tests that are specific to academic fields (e.g., Economic Records Examination) seem to be more popular recently.

There is a wide gap in the degree of prevalence among four direct-incomparable (D-I) assessment methods which are being listed below D-C ones. In contrast with the other two methods, both student grades and (the performance in) senior year research are treated as learning outcomes in more than 60% of faculties. Most notably, the proportion of the faculties that developed original examinations for outcomes assessment is only 10.6%.

The last type of assessment methods, the remaining four in Fig. 1, is indirect-incomparable (I-I). While this type of assessment generally consists of various methods that focus on subjective perceptions gained through such as interviews and focus groups, what is looked at here is the data from surveys administered to four types of respondents. The figure shows that people inside the university are more frequently targeted than those outside. However, what can be said regardless of respondents' traits is that the number of faculty adopting opinion surveys has been increasing remarkably since 2000.

Use of assessment results in self-review

Even if multiple types of outcomes assessment are implemented, it will not lead to educational improvement unless the results of assessment are accounted for during the process of self-review. Figure 2 reflects the actual reality of the situation on this, with reference to the four areas of self-review. The percentages in this figure equal the proportion of faculties which uses at least a single method included in each type of outcomes assessment.

In self-reviews concerning the educational purpose of faculty, I-C assessment is most frequently used (44%). The other types follow in order of I-I, D-I, D-C. The same pattern can be seen in the case for curriculum, i.e., selection and arrangement of courses, and for contents, i.e., what was taught in each course. In contrast, I-C assessment is given less consideration than incomparable assessments in self-reviews of educational methods.

Figure 2 suggests that the degree of adoption varies depending on the type of assessment and review area. Then, one can ask whether the adoption of outcomes assessment has an impact on the improvement of education. To answer this question, logistic regression analysis was employed. The variables used in the analysis are shown in Table 2.

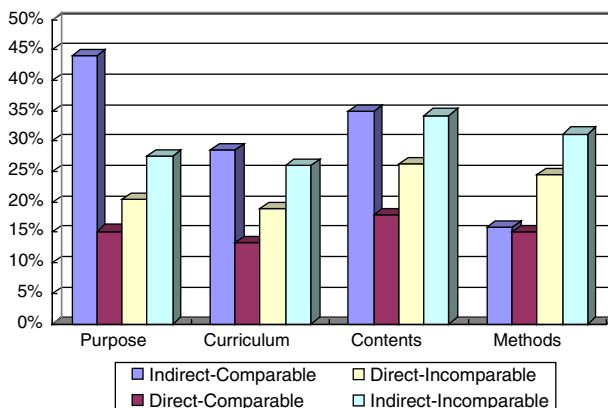


Fig. 2 Types of assessment methods used in self-reviews

Table 2 Variables used in logistic regression analysis*Dependent variables*

Improvements of education are perceived: yes = 1, no = 0

Independent variables

Traits of faculty

Sector of university: national = 1, public and private = 0

Relation to vocations: “very strong” = 5 to “very weak” = 1

Admission difficulty: t-score of admission examination

Traits of self-review

Internal demands: “very strong” = 5 to “very weak” = 1

External demands: “very strong” = 5 to “very weak” = 1

Permanent committee for self-review: yes = 1, no = 0

Degrees of discussion in the committee: “frequent” = 4 to “never” = 1

Adoption of outcomes assessment

Indirect-comparable assessment: yes = 1, no = 0

Direct-comparable assessment: yes = 1, no = 0

Direct-incomparable assessment: yes = 1, no = 0

Indirect-Incomparable assessment: yes = 1, no = 0

The dependent variables are whether the improvement of education is perceived in each area. The proportion of faculty which perceived improvement is 69.7% in purpose, 76.5% in curriculum, 82.3% in contents, and 76.9% in methods. The independent variables are divided into three categories. Traits of faculty comprise sector, vocational relevance, and difficulty of entrance of respondent faculty. Traits of self-review comprise purposes of evaluation, existence of permanent committee for self-review, and the degree of discussion in the committee. Adoption of outcomes assessment is defined as yes or no in accordance with the answers summarized in Fig. 2.

Table 3 indicates the results of the analysis. Throughout the four areas, the “Internal demands” predictor has definite influence. It means that improvements are more likely to be perceived, when the faculty has a clearer awareness that the self-review is taking place for the purpose of improving education. Again, though the significance levels vary, being a national university, having a higher recognition of external demands and having frequent discussions by the self-review committee also have a positive effect in some areas.

Looking at the results concerning applications of assessments, one can find only limited effects. Adopting I-C assessment significantly raised the possibility of acquiring affirmative perception of improvements in purpose and curriculum. I-I assessment was adequate for improvements in curriculum and educational methods. No significant relationships were found for D-C and D-I assessments.

What has been clarified and what has not

Discussions

Several points will be discussed by combining the reforms explained in the second section and the results of the analysis in the third section.

Table 3 Determinants of improvement of undergraduate education

	Purpose	Curriculum	Contents	Methods
National university	1.595 ⁺	2.365**	1.238	1.733 ⁺
Relation to vocations	1.273	1.086	0.871	0.781
Admission difficulty	1.013	1.004	1.016	0.993
Internal demands	1.825**	1.659**	2.694***	2.545***
External demands	1.191	1.447**	1.285 ⁺	1.298 ⁺
Permanent committee	0.954	1.147	0.890	1.274
Degrees of discussion	1.396 ⁺	0.952	1.144	1.461*
Indirect-comparable	1.688*	2.216*	1.523	1.420
Direct-comparable	1.073	1.548	1.619	1.619
Direct-incomparable	0.866	0.536	0.801	1.036
Indirect-incomparable	1.367	1.901 ⁺	1.219	1.800 ⁺
Model chi-square (<i>df</i>)	66.612 (11)***	61.638 (11)***	58.368 (11)***	89.957 (11)***
Negelkerke R^2	.191	.190	.198	.270
% Correct prediction	76.4	79.0	85.4	81.0

Statistics for predictors are odds ratio

*** $P < .001$, ** $P < .010$, * $P < .050$, ⁺ $P < .100$

First of all, an obvious influence from external demand on outcomes assessment was seen in its degree of prevalence. As shown in Fig. 1, the commencement of evaluation by NIAD was more influential than the regulation seeking voluntary self-review. This was especially noticeable in indirect-incomparable methods, which could be explained by the fact that NIAD required subjective opinions about learning outcomes in its evaluation. It has turned out to be the case in Japan, as well as in the U.S. (Lubinescu et al. 2001) and Europe (Adam 2008), that external demands urge HEIs to carry out outcomes assessment.

As a second point, it was shown that the results of assessments had not been utilized in self-reviews in nearly half the faculties which had implemented the assessments themselves. According to the aforementioned definition of assessment use, the percentage of faculties using a type of assessment is no greater than that of the most used method of the same assessment type. Therefore, it can be said that D-C assessment was used in at least 50.3% of faculties since the percentage of “national examination” use was the most used D-C assessment method at 50.3%. In the same manner, I-C assessment was used in at least 85.4% of the faculties, D-I assessment in at least 73.2%, and I-I assessment in at least 67.8%.

However, even when selecting an area with the highest self-review adoption rate, D-C assessment was only used in 17.9% of faculties for contents-reviews. Similarly, I-C was only used in 44% of faculties for purpose-reviews, D-I was only used in 26.3% of faculties for content-reviews, and I-I was only used in 34.3% of faculties for contents-reviews. These results ironically suggest that outcomes assessment which began through external demands were not always useful for internal purposes. In other words, assessments were clearly being carried out, but may not have been supporting the improvement of education.

Finally, the results of logistic regression analysis also revealed the situation in which outcomes assessments were being utilized inadequately. The influence of outcomes assessment adoption on educational improvement was limited, while traits of faculty and self-review had explicit effects. The results showed that national universities which had

been exposed to NIAD's trial evaluation advanced educational improvement. Moreover, the importance of purposeful self-review, whether as a result of internal or external demands, was made apparent.

These findings reinforce the hypothesis that outcomes assessment initially motivated by external demands do not necessarily play an active role in self-review, and consequently, do not lead to the progress of undergraduate education. But it is not a reasonable decision to give up outcomes assessment, because the demand for it would not be expected to diminish. The experiment to develop AHELO by OECD is a plain evidence supporting such observation. Rather than waiting for that "outcomes assessment trend" is over, considering worthwhile ways to use outcomes assessment is more constructive.

Keys to further research

One of the ways is that eliciting the relationship between learning outcomes and activities of education/learning prior to self-review. The absence of the elicitation is often indicated as a reason of misconnection between outcomes assessment and improvement. In the general discussion of evaluation, evaluations using causal mechanism between activities and outcomes are called such as "Program Theory Evaluation", "Theory-Based Evaluation" and "Theory-Driven Evaluation" (Rogers et al. 2000). Although the term "Program Theory" is not familiar to researches in higher education, similar things like matrixes explaining the relation between learning outcomes and individual courses surely exist (e.g., Diamond 1998). Investing how much those "Theories" elicited contribute to beneficial outcomes assessment is a theme worth to research.

Another way is rather conceptual one comparing to the aforementioned practical way. That is changing the idea on "good university education". Suppose according to Lee (2003), an evaluation, as well as self-review, is the process of determining merit or value based on the information gained through assessments. And learning outcome is just an example of the information. The kind of information referred in evaluation depends on the idea of evaluators on "good university education", and of course, the ideas vary evaluator by evaluator, plus, situation by situation (Brennan and Tarla 2000). While it is a natural idea to refine education based on the information on student learning outcomes as any educational activity aims to produce learning, it also can not be denied that the goodness of university education have been sometimes defined by academic perspectives instead of educational ones. If some reasons of restricted profits of outcomes assessment in self-review lies in the concept of faculties on good education, it is probable that research on it will partly reveal the causes of the problem.

References

- Adam, S. (2008). Learning outcomes current development in Europe: Update on Issues and the applications of learning outcomes associated with the bologna process. Paper presented at Bologna Seminar at Heriot-Watt University, Edinburgh, Scotland on 21–22 Feb 2008. <http://www.ond.vlaanderen.be/hogeronderwijs/bologna/BolognaSeminars/Edinburgh2008.htm> (Accessed on 30 June 2009).
- Amano, I., & Poole, G. S. (2005). The Japanese university in crisis. *Higher Education*, 50, 685–711.
- Barr, R. B., & Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change, November/December*, 13–25.
- Brennan, J., & Shah, T. (2000). *Managing quality in higher education: An international perspective on institutional assessment and change*. Buckingham: OECD, SRHE and Open University Press.
- Diamond, R. M. (1998). *Designing and assessing courses and curricula*. San Francisco: Jossey-Bass.

- Dill, D. D. (2000). Designing academic audit: Lessons learned in Europe and Asia. *Quality in Higher Education*, 6(3), 187–207.
- Ewell, P. E. (1997). Strengthening assessment for academic quality improvement. In M. W. Peterson, D. D. Dill, & M. A. Lisa and associates, *Planning and management for a changing environment: A handbook on redesigning postsecondary institutions* (pp. 360–381). San Francisco: Jossey-Bass.
- Huba, M. E., & Freed, J. E. (2000). *Learner-centered assessment on college campuses*. Boston: Allyn and Bacon.
- Kitamura, K. (1997). Policy issue in Japanese higher education. *Higher Education*, 34, 141–150.
- Kushimoto, T. (2004). Outcome assessment of liberal education in Japan. *Japanese Journal of Higher Education Research*, 7, 137–155.
- Kushimoto, T. (2006). The present state and challenges of program evaluation in university education: Toward the establishment of formative evaluation based on learning outcomes. *Daigaku-Ronsyu (Research in Higher Education)*, 37, 265–276.
- Lee, W. Y. (2003). Preface. In W. Y. Lee (Ed.), *Assessment and program evaluation* (2nd ed.). Boston: Pearson Custom Publishing.
- López, C. L. (2004). A decade of assessing student learning: What we have learned and what is next. In P. Hernon & R. E. Dugan (Eds.), *Outcomes assessment in higher education*. Westport, CT: Libraries Unlimited.
- Lubinescu, E. S., et al. (2001). Two continuums collide: Accreditation and assessment. *New Directions for Higher Education*, 113(Spring), 5–17.
- Ministry of Education. (2008). Daigaku ni okeru Kyouiku-Naiyou-tou no Kaikaku-Jyukyoku ni tuite (Progress of University Education Reform). http://www.mext.go.jp/b_menu/houdou/20/06/08061617.htm (Accessed on 3 September 2008).
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe, 1986–1988. *European Journal of Education*, 23(Nos1/2), 7–23.
- Palomba, C. A., & Banta, T. W. (Eds.). (2001). *Assessing student competence in accredited disciplines*. Sterling VA: Stylus.
- Peterson, M., Einarson, M., Augustine, C., & Vaughan, D. (1999). *Institutional support for student assessment: Methodology and results of a national survey*. Stanford: National Center for Postsecondary Improvement.
- Research Institute for Higher Education Hiroshima University. (1999). Higher education and market-oriented reform.
- Rogers, P. J., et al. (Eds.). (2000). *Program theory in evaluation: Challenges and opportunities. New directions for evaluation 87(fall)*. San Francisco: Jossey-Bass.
- University Council (Daigaku-Shingikai). (1998). A vision for universities in the 21st century and reform measures.
- Yonezawa, A. (2002). The new quality assurance system for Japanese education: Its social background, tasks and future. *The Japanese Journal of University Evaluation of National Institution for Academic Degrees*, 2, 21–34.