

## **Session 2-2**

### **Rethink of Peer Review Process in Quality Assurance of Higher Education from Analysis of Appeal Report in Taiwan**

**Karen Hui-Jung Chen**

Higher Education Evaluation and Accreditation Council of Taiwan

**Angela Yung-Chi Hou**

Fu Jen Catholic University

**Chia-Chun Ho**

Higher Education Evaluation and Accreditation Council of Taiwan

**Yung-Yung Chang**

Higher Education Evaluation and Accreditation Council of Taiwan

**Li-Yun Wang**

Higher Education Evaluation and Accreditation Council of Taiwan

**Hui-Ling Pan**

Tam Kang University

#### **Abstract**

Peer review has been a cornerstone of quality assurance of higher education. It has recently draw criticism due to reviewers' judgments might be not based on actual achievements. In order to enable the assessed programs to be against the accreditation decisions, an objection and appeal system has been developed in Higher Education Evaluation & Accreditation Council of Taiwan (HEEACT). A program could file an objection if the evaluation result is inconsistent with fact or violating procedures after the draft of the site visit report is completed. The purpose of this empirical study is to explore the major reasons for the objections made by institutions through appeal reports as well as examined the validity and reliability of the peer review systems. Content analysis was conducted to analyze 5039 items of 642 objection reports from 65 universities. Results showed that the objection rate of each program is decreasing each year. The institutions disagreed with the draft of on-site visit report for the reasons: *inconsistent with the fact, and asking modification of the rhetoric on the reports*. Concerning the validity of peer review, we compared the views of university

staff and reviewers and found differences existed in the following areas: *lacking context analysis of the programs, invalid information, unreliable information, unjustified conclusion, partial reports, and not based on the same criteria for making judgments*. Concerning the reliability of peer review, it showed that reviewers might not be based on the same criteria to write evaluation reports and make judgments on three issues. Also the quality of peer review could be influenced by the context of the whole educational system: *limited size of educational system, and the close relationship among the reviewers*. These factors increase the complexity of the peer review process.

## **Introduction**

Peer review has been a major element of quality assurance of higher education (Harvey, 2002). Having experts with deeply understanding the field being evaluated, peer review was able to make judgment and comments for the external evaluation. However, it has recently drawn criticism due to reviewers' judgments might be not based on actual achievements. The prior experience of peer reviewers tends to influence the decisions. (Adamson & Flodstrom, 2012; Huisman & Currie, 2004; Silva, Reich & Gallegos, 1997; Kristoffersen, 2012). In order to enable the assessed programs to be against the accreditation decisions, an objection and appeal system has been developed since Higher Education Evaluation & Accreditation Council of Taiwan (HEEACT) started its first cycle evaluation in 2006. According to HEEACT, a program could file an objection if the evaluation result is inconsistent with fact or violating procedures after the draft of the site visit report is completed. The aim of this study is to explore the major reasons for the objections made by institutions through appeal reports as well as examine the validity and reliability of the peer review systems.

## **Theoretical Background**

### ***Validity and Reliability of Peer Review***

Peer review has been a widely applied method in the quality assurance of higher education. Peer reviewers conducted mixed methods with both quantitative and qualitative approaches in the on-site visit to check the self-evaluation reports. They gave comments and wrote evaluation reports according to their observation and judgments. The primary focus of peer reviewers is to capture the authentic practices of teaching, learning, and administration of the programs during a short on-site visit period. It has been accepted that peer reviewers are the data collection instrument in the peer review process. They tried to map the whole picture of educational programs and check the self-evaluation reports by conducting observations, in-depth interviews of administrators, staff, teachers, and students in the university settings. The whole

on-site process for peer reviewers is just like the researchers of qualitative studies tried to make a picture of peoples' experiences by making research plan, collecting data, interpreting data, making inferences about the findings, and drawing conclusion.

When we analyzed the whole process of on-site visit of peer review, we could found that it relies on an assumption that the reviewers are objective and neutral. Peer reviewers could objectively find out the real situations of programs by multiple data collecting and interpretation process, and making judgments independently without bias but with their academic professionals. With this assumption, the challenges of conducting fair peer reviews are to maximize the validity of peer review by focusing on the methodologies and procedures, and do not need to consider the personal or emotional factors of individual reviewers. If we could identify the threats to validity and have good control on these, the validity of peer reviews could be improved (Greene, 2011; House, 2011). For example, data collections could be distorted by conscious or unconscious way during the collecting process. Strategies could be conducted for enhancing the validity of the data collection, such as triangulation: multiple methods for obtaining data, multiple data sources, and more than one researcher to collect and interpret data (Webb, Campbell, Schwartz, & Sechrest, 1966; Denzin, 1978). Maxwell (1992) mentioned five types of validity: (1) descriptive validity: factual accuracy of the account as documented by the researcher; (2) interpreting validity: the extent to which an interpretation of the account represents an understanding of the perspective of the members' words and actions; (3) theoretical validity; (4) generalizability; (5) evaluative validity.

On the other hand, this assumption of neutrality and objectivity has been challenged by Greene (2011) and other researchers. Greene proposed that the peer reviewers are neither neutral nor objective. They conducted evaluations in a situated context and inevitably brought their own socio-cultural history and beliefs into the process. It is an interpretive inquiry process and inevitably imbued with their own lenses in the review and interpreting processes. Therefore it was suggested that the interpretive validity is not how to generate objective inferences correspond to the real evaluated programs, but rather a matter of generating meaningful inferences according to the evidence, and turned those causal inferences to evaluation results.

There is another challenge in the interpretive peer review process. The focus of the on-site visit is the outcomes, which is seemed as isolated phenomena that can be studied independently of process and interactions with interpretive view of the university settings. However, the process is contextual and dynamic. The results could be changed dramatically different by changes of the reviewers with different knowledge, attitudes, beliefs, and experiences. The validity of peer review is suggested as arguments, constructed rhetorically from evidence and shared

conversations (Greene, 2011; Lather, 1993; Kane 2006). Peer Reviewers are required to align, integrate, and interpret evidence into convincing arguments.

Validity of the arguments of peer reviewers is not to find out the objective truth but understand university actions in the context. In this interpretive view, validity for inferences from an evaluation is to investigate actions in three dimensions: context, relation, and values. Each program is situated in a particular context, with special social, cultural, and economic characteristics. It should be noticed that those factors interwoven together and affect the whole context of the program. As to the relation dimension, the peer reviewers are having an up-close, on-site and direct observation and conversation with multiple stakeholders. They interpreted data and made inferences in the ongoing and active relationship with multiple stakeholders (Greene, 2005). The last, peer reviewers made judgments in the evaluation and thus values engaged. Each criterion for the judgments reflects different values, and also different stakeholder interests. Validity of peer review in the interpretive perspective is to make persuasive communications with empirical evidence rather than objective judgments.

It has been added interpretive perspective in the on-site visit, considering the importance of context, complexity, and contingency in understandings in the incidence, cause and meaningfulness of the observed phenomena. The reviewers are asked to make inferences correspond to the real university context, rather than generating inferences to universal phenomena and criteria. However, it has been pointed that it is very difficult for reviewers to make judgments about the outcomes/effect of an educational program. The reviewers only looked the outcomes as isolatable phenomena that could be studied independently of process and interactions and ignored the contextual and dynamic factors. Therefore it has been proposed three kinds of threats from: the contextual dimensions of human action, the relational and dialogic facets of evaluation, and the value engaged.

In this study, we defined threats of validity as correlation between review decision and perspectives of university staff, and threats of reliability as agreements between reviewers. The threats of validity and reliability of peer review in the on-site visit in investigated according to the empirical data of the objection reports.

### ***File Objections to the Results of Peer Review***

Quality of higher education is ensured by internal and external quality assurance. A trend in many countries is to conduct self-evaluation as internal quality assurance and on-site visit as external quality assurance to ensure continuously improve the quality of higher education. The evaluation results rely on the methods of peer review, with people who are able to make judgments and recommendations for improvement (Vanhoof & Petegem, 2007). However, it has draw attention that the reviewers'

judgment might be influenced by their prior experiences (Harvey, 2002). Therefore the appeal systems are set up in the quality assurance of higher education in many countries to let university have chances to say something about themselves and make balances between the perspectives of peer reviewers and university staff.

Evaluation of higher education in Taiwan applied the accreditation module and an appeal process is included in the evaluation system design. The first cycle of program evaluation was launched by Higher Education Evaluation and Accreditation of Council in Taiwan (HEEACT) in 2006. HEEACT adopted accreditation model to conduct the program evaluation, using self-evaluation and on-site visits to understand the quality of universities. Peer reviewers identify the gaps between the desired and the present situations through self-evaluation and on-site visit. After program evaluation, each program will be granted one of the accreditation statuses: accredited, accredited conditionally, or denial. In order to protect the rights of universities, universities in Taiwan could file an objection or appeal if they did not agree the evaluation results. An objection is raised after on-site visits. After the on-site visit, HEEACT will send a copy to the institutions evaluated after the draft of the on-site visit report is completed. If they regard the comments in the report are inconsistent with fact, or think reviewers violated certain procedures during the on-site visits, they can file an objection to make sure the fairness of this evaluation. In response to the objections from the institutions evaluated, HEEACT will invite the on-site visit team to review the opinions on the objections and make sensible decisions. A report will be sent to the university.

### ***Research Questions***

The aim of this study is to explore the major reasons for the objections made by institutions through appeal reports as well as examined the validity and reliability of the peer review systems. The two research questions are:

1. What are the appealing rates and accredited rate in the first cycle of program accreditation in Taiwan from 2006 to 2009?
2. What are the major reasons for the objections made by institutions?
3. What are the differences between the perspectives of university staff and peer reviewers? How did these differences affect the validity and reliability of the peer review systems?

### **Method**

The first cycle of program accreditation of higher education was conducted by HEEACT from 2006 to 2010 in Taiwan. Due to the program evaluated in 2010 were police and military program, the evaluation results were not opened to the public. This

empirical study analyzed the appeal reports from 2006 to 2009 in Taiwan. It includes the objection reports of 672 programs and totally 5039 appeal items in these reports. All appeal reports could be downloaded from the website of HEEACT. For this investigation, only the appeal reports of national and private universities were analyzed, not including the reports of the military institutions. Data analysis in this study consisted of two stages. Firstly, the appeal reports from different programs were analyzed by years to understand the changes of the appealing rates and accreditation passing rate. Secondly, we conducted content analysis of the appeal reports to investigate the perspectives of institution staff and peer reviewers respectively.

## **Results and Discussion**

### ***Appealing and Accreditation-Passing Rate Analyzed by Years***

In 2006 to 2009, a total of 1818 programs have been evaluated in Taiwan. The proportion of the appealed programs was approximately reducing every year, from 61% in the June of 2007 to 30% in the December of 2009. There were 1818 programs being evaluated in Taiwan, 82% of them got accredited, 16% conditional accredited, and 2% denial. It showed that the accredited rate was approximately increasing every year, from 66% in the June of 2007 to 90% in the December of 2009. The rate of conditional accredited was reducing from 23% in the June of 2006 to 10% in the Dec of 2009.

### ***Responses from institutions***

It was revealed that the ratio of submitting appeal reports to the program being evaluated is decreasing each year (from 59% to 25%), however the average items for each program is increasing every year (from 5.3 to 8.0 items). The institutions disagreed with the draft of the site visit report for several reasons. First of all, 80% of the appeals considered the reports were inconsistent with the fact. They provided more information and explained how the situations really were. Secondly, 9% of the appeals agreed with the reviewers' recommendations for improvement, but they requested to modify the rhetoric on the reports, such as changing the description from "strong" to "some". The third reason for national and private institutions is different. The national institutions concerned more about how the data collected by the reviewers, such as reviewers making judgments by their prior perceptions of the institution. On the other hand, the private institutions concerned more about the specific context of the institution. They wondered the reviewers applied the universal standards to evaluate each institution.

### ***Responses from Peer Reviewers***

From the perspective of the reviewers, it showed that most reviewers seem to be keep their opinions unchanged. However, the percentage of upholding the original report decrease (from 76% in 2006 to 63% in 2009), and the percentage of revising the report also decreased (from 22% to 15%), while the percentage of upholding parts of the report is increasing (from 2% and 15%).

### ***Validity of Peer Review***

Concerning the validity of peer review, we compared the views of university staff and reviewers and found differences existed in the following areas: lacking context analysis of the programs, invalid information, unreliable information, unjustified conclusion, partial reports, and not based on the same criteria for making judgments.

#### ***1. Context Analysis of the Programs***

The evaluation should describe the program's background and setting (Stufflebeam, 2000). From the content analysis of the appeal report, the university staff proposed that the special context of their programs should be considered. The results revealed that 38 items (0.8%) related to this issue. The percentage is very low in this category as compared to other categories. We further analyzed how university staff and reviewers perceived this issue by content analysis. The items (36 items) proposed by university staff are much more than by reviewers (2 items). From the perspectives of university staff, they proposed the reviewers should consider the special contexts of their programs: special or broader definitions of program goals (31%), newly established programs (14%), program goals changed (11%), inter-discipline programs (9%), and special nature of disciplines (6%). While from the perspectives of reviewers, only two appealing items related to this issue. From the above, we found that the views of university staff were quite different from reviewers and this might cause the significant discrepancy between self-evaluation and on-site visit.

#### ***2. Impartial Reporting***

The reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings (Stufflebeam, 2000). Results of content analysis showed that only 10 items in this category (0.2%). The low percentage indicates that the on-site visit reports are impartial as considered by both the university staff and reviewers. However, it is important to identify the bias causing the impartial reports.

From the university staff sides, they considered bias reviews coming from: reviewers' prejudices of the programs, and reviewers' violation of the principle of confidentiality. On the other hand, few reviewers replied with emotional wording and those were classified into this category.

### **3. *Focusing on the Same Issue***

The focused issue of university staff and reviewers during the back-and-forth appeal process should be the same in order to keep the arguments properly. After comparing the comments of on-site visit report, universities' appeals, and the reviewers' replies, we found that 106 items (2%) were classified into this category. We further analyzed the reviewers' replies from the reports and found that four actions of universities might cause this happened: giving irrelevant answers (71 items, 67%), providing data of improvements after on-site visit (19 items, 18%), giving insufficient or not specific answers (17 items, 16%), misinterpreting the on-site visit reports (2 items, 2%), and self-contradictory arguments proposed by universities (1 items, 1%).

### **4. *Valid and Reliable Information***

The reviewers were expected to use several methods to gather information during the on-site visit to verify and validate the self-evaluation report submitted by department staff. The methods include interviews (students, faculty, and administrators), large group meetings, reading documents, classroom and infrastructure observation. Through the triangulation process, the reviewers could get a whole picture of the program. Stufflebeam (2000) also pointed out that the gathering procedure should be chosen or developed, and then information obtained is sufficiently reliable for the intended use. Results of content analysis showed that 363 items (7%) were classified into this category of valid and reliable information. Four reasons were identified for causing invalid and unreliable information, including over emphasis on interview results (304 items, 6%), fail to clarify the uncertainty during on-site visit (30 items, 0.6%), inadequate data citation (10 items, 0.2%), and university providing improvement data conducted after on-site visit (19 items, 0.4%). The reviewers emphasized on the information gathered from student interviews. The information needs to be triangulated through checking with other data. If the replies could be provide more evidence from other sources, the information will be more reliable.

## **5. *Justified Conclusion***

The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them (Stufflebeam, 2000). In this study, we found that the university staff perceived the reviewers' suggestions as unjustified conclusion for two main reasons: the suggestions were difficult to achieve, and the suggestions were vague and not specific. Although the percentage is not high (31 items, 0.6%), the phenomena is important to be analyzed in order to make improvements of the evaluation process. The first reason shows that some of the universities considered that the suggestions were difficult to achieve (24 items, 0.5%, S5b). According to the results of content analysis, the main reasons were the suggestions were related to the changes of the whole university system, or lots of funding needed to make improvements. For example, the reviewers asked a newly established department to be internationalization, which needs lots of funding and it is difficult to be achieved for a new department. And the reviewers suggested another department to build up the flexible salary system to include world class experts as faculty. The universities complained that it is very difficult for a department to make such a big change of the whole system without university support. The second reason was that universities considered that reviewers' suggestions were vague and not specific (7 items, 0.1%, S5a). All of these come from the wordings of on-site visit reports and the university staff felt it was very difficult to catch the real meanings of reviewers' comments. In addition, we analyzed the appeal items rejected by the reviewers (2669 items, 53%) and found that the reviewers' replies could be classified into two types: replies without explanations (749 items, 15%) and replies with detailed interpretations (1920 items, 38%).

### ***Reliability of Peer Review***

It was found the reviewers used inconsistent standards to make judgments and recommendations. Some of the reviewers refused to accept the data provided after on-site visit. They only took the data shown on the days of on-site visit, even though it is the responsibility of reviewers to ask questions if they were not sure of the situation. On the other hand, some reviewers applied low standards to accept the appeal as long as the institutions claimed they have an improvement plan. In addition, the reviewers applied different criteria to judge whether the universities provide valid data or not. Thirdly, the way reviewer writing reports are in very different way. Most reviewers (71%) explained the reasons why they did not accept the appeals, while other reviewers did not give any reasons.

### ***Quality of Peer Review***

The application of peer review is influenced by the context of the educational system. As the limited size of educational system in Taiwan, it is difficult to involve large and various panel members in the same discipline for peer reviews. It might happen that the reviewers conduct mutual evaluation. The reviewers in small states have to face the dilemma and the pressure from their colleague from other institutions. If the funding is related to the evaluation results, the pressure is even higher. The complexity of peer review is increasing as the higher education system is smaller.

### **Concluding Remarks**

Peer review is the cornerstone to constitute sound quality assurance procedures. The validity and reliability of peer reviews is being challenged. The skills and knowledge of reviewers, and the variety of reviewers involved needs to be seriously reviewed. Results showed that the objection rate of each program is decreasing each year. The institutions disagreed with the draft of on-site visit report for the reasons: inconsistent with the fact, and asking modification of the rhetoric on the reports. Concerning the validity of peer review, we compared the views of university staff and reviewers and found differences existed in the following areas: lacking context analysis of the programs, invalid information, unreliable information, unjustified conclusion, partial reports, and not based on the same criteria for making judgments. Concerning the reliability of peer review, it showed that reviewers might not based on the same criteria to write evaluation report and make judgments on three issues. Also the quality of peer review could be influenced by the context of the whole educational system: limited size of educational system, and the close relationship among the reviewers. These factors increase the complexity of peer review process. Analysis of the content of the reports, the gap between university staff and peer reviewers could be understood. It helps to improve the validity and reliability of peer review.

### **References**

- Adamson, L. & Flodstrom, A. (2012). Political meddling in quality assurance spells chaos. University World News, Issue 00247, Nov. 14, 2012. Retrieved from <http://www.universityworldnews.com/article.php?story=20121107150730970> on Nov. 14, 2012.
- Christie, C. A. (2003). What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory. *New Directions of Evaluation*.
- Denzin, N. (1978). *Sociological methods*. New York: McGraw Hill.

- Greene J. C. (2011). The construct (ion) of validity as argument. In H. T. Chen, S. I. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation. Theory and Practice. New Directions for Evaluation*, 130, 81-91.
- Harvey, L (2002). Evaluation for What? *Teaching in Higher Education*, 7(3), 245-263.
- House, E. R. (2011). Conflict of interest and Campbellian validity. n H. T. Chen, S. I. Donaldson, & M. M. Mark (Eds.), *Advancing validity in outcome evaluation. Theory and Practice. New Directions for Evaluation*, 130, 69-80.
- Kemenade, E. V. & Hardjono, T. W. (2010) A critique of the use of self-evaluation in a compulsory accreditation system. *Quality in Higher Education*, 16 (3), 257-268.
- Kristoffersen, D. (2012). Evolutionary approaches to peer review and future directions. *Evaluation in Higher Education*, 6(1), 1-14.
- Silva, M., Reich, R. & Gallegos, G. (1997). Effects of external quality evaluation in Chile: a preliminary study. *Quality in Higher Education*, 3(1), 27-36.
- Vanhoof, J, & Petegem, P. V. (2007). Matching internal and external evaluation in an era of accountability and school development: lessons from a Flemish perspective. *Studies in Educational Evaluation*, 33 (2), 101–119.
- Vroeijenstijn, T. (2011). Internal and external quality assurance – why are they two sides of the same coin? INQAAHE 2011 conference.
- Webb, E. J., Campbell, D. T., Schwartz, R., & Sechrest, L. (1966). Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally.
- Zhang, J. (2012). Thoughts on internal and external quality assurance. *Educational Research and Review*, 7(8), 196-200.