NUMBER'S CONSEQUENCES – AN EXPERIMENT AND A DISCUSSION IN STANDARDIZATION OF TEACHING EFFECTIVENESS EVALUATIONS BASED COURSE PERFORMANCE STATISTICS

Hong Wu[1]

Abstract — *It is both interesting and worthwhile to experiment new methods for online course's effectiveness evaluations. The current study is an attempt for such an unconventional method. A survey of a number of sampled online courses is conducted on these courses' performance statistics, as their numbers of registered course documents. The survey analysis has examined these courses' performance statistics and established an empirical performance summary from each course. Further analysis compared on these performance summaries each other and mapped correlations between these performance summaries. As a result, a measurement standard for effectiveness evaluations for these online courses is established, empirically. The current study has also discussed the validity and reliability issues in standardization of teaching effectiveness evaluations based on courses' performance statistics.*

Key words: Standards for teaching effectiveness, course data survey, performance statistics, evaluations for online teaching and learning.

IS FINAL EXAM THE ONLY WAY TO CHECK?

The final exam is a most common method to check the effectiveness of a teaching course. The method has been used several thousands years and it still used widely in most of academic institutions. No doubt many brilliant learners became experts through this check method.

However, like other methods, there are both advantages and disadvantages of this method as well. Whenever a final exam is the only way to check the effectiveness of a teaching course, through a student's exam answer sheet, it would easily become an exam game. As we all probably know, a student would sometimes focused on exam details, rather than the syllabus that have been taught through. As we all probably remember our own student age, that how easy for us to prepare intensively and instantly for the final exam only two weeks beforehand, but still received good marks, though we might have escaped from lectures and reading for soccer games many times under the course semester.

To avoid such cases, prosess evaluation method is introduced and one final exam is devided into many small, independent tests throughout the semester. It is perhaps a well balansed method to check the effectiveness of a teaching course. However, it requires also more resources and attention, both from teachers and students, to prepare and accomplish theses tests, so the prosess itself might overrun the lecture's content.

Both menitoned methods are relatively direct evaluation and both seem to be either too much randomly or too costly. One alternative is looking for some indirect evaluation, such as performance based evaluation to compensate or support the final exam method.

What is a performance based evaluation? We can think about the similar case for TQM (Total Quality Management) philosophy and its application in the manufacturing industry. The TQM principle is based on the prosess and performance evaluations, rather than final control and inspections. It says if every step of a prosess, or everyone performs their tasks well, the final product will then be good.

Similarily, a performance based evaluation for teaching class considers every class related activities as a part of learning prosess. Say, how many questions a student has asked during the lecture? How many pages a student has read for this chapter? How often a student discusses with other classmates? How many hours a student spend for reading, etc. As a Chinese motto says, if you have traveled thousands of miles, you are surely knowledgeable.

Compared with traditional classroom course, online course offers a great advantage in such performance based evaluation methods. Since online communication must be undertaken electronically, so every activity, questions, chat, tasks, etc. will be restored in database, automatically. This will make data analysis easily. However, there is a need for discussions regarding the measurement standards for online course, so an evaluation will have its real meaning.

The current study has sampled few online courses to illustrate this issue.

SAMPLED ONLINE COURSES

Our university college's has designed an online introduction course for college teachers and professors, and this course has been successfully conducted several times since 2000. The current study has sampled 3 particular online course cases, conducted at April 2001, June 2001 and April 2002, respectively.

The course is designed as a 3-weeks intensive, but an introduction level. The target group is college teachers and professors, and the intention is motivating and stimulating their interests in online technology for their own courses.

[1] Hong Wu, Østfold University College, Norway, N-1705   hong.wu@hiof.no

The majortity part of this course content is practicing participants' online skills. Therefore, the course is named as "Online Teaching – Introduction and Basic Practice" (OTIP), or "Nettbasert Undervisning" (NBU) in Norwegian. The English version of the course is designed for foreign professors and the course was actually conducted twice outside of Norway, once in June 2001 in China (sampled for the current study), and another time in September in Iceland. The Norwegian version of the course has been conducted 4 times already and there were nearly 100 participants entered the course.

To symbolize these 3 sampled online course cases, the following course label is used for further analysis, as NBU2 for the course conducted at April 2001, OTIP1 for June 2001 and NBU4 April 2002.

ONLINE PERFORMANCE – NEW DEFINITIONS

What is online performance? Why new definitions? It is important to clarify these questions before detailed analysis and comparisons. As we all know, online activities are different from a traditional class room teaching environment, so ways of class performance have to be different, so do the definitions.

For example, a good student in a traditional class room can be noticed as a one whom is actively answering a teacher's questions, or actively engaging in the class discussions, or accomplishing excercises and assignments, both in class and after class. The majority part of these activities would be oral based and some would be written based. Therefore, a student with A personality would probably be reaching ahead easier than another with B personality. Also, we have observed a good student by impression, because the most part of class performance was undertaken by oral expression.

Online class situations would probably be different, since our current courseware is written based. A good online student could still actively answer questions, engaging in the class discussions, or accomplishing assignments. However, the majority part of such activities are written based and this change provides an opening, not only for A personality, but also for B personality students. Since everything is in written form and restored in the database, we could extract or track up a good online student by record, because the most of class performance was undertaken by writing.

Since online class performance is so different from the traditional class room teaching, there is a need for new definitions. The great challenge is identifying new parameters in a new environment. On the other hand, the great advantage of online teaching is that everything is recorded in the database, automatically. This makes data collection and data analysis easy.

Compared with traditional class room teaching, there are few remarks for online teaching that should be noticed in distinction between these two approaches. Online teaching environments are rather:

- 24 hours continually
- Less spontanous, more basic actions
- Well prepared thinking and expressions.
- Many questions, but also urgent for answers
- Easy to misunderstand, but difficult to explain online
- More students have more accesses for more questions

These remarks have made a clear distinction between an online teaching environment and a traditional class room teaching environment. The distinction can also be a basic circumstance for the new definitions of online class performance. We can generally believe that in an online teaching environment, a student should:

- Be flexible, but active and initiative
- Be patient, but systematical and self constrcutive
- Think clearly and express yourself clearly
- Question many instead of many questions
- Be considerable, write clearly and writr in details
- Seeking for answers, but not waiting for answers

Hence, online class performance and its measurement should be based on the mentioned criteria. In practice, it is quantity of the mentioned activities that counts. This means we need to establish a meaurable standard or a group of parameters that indicates online class performance.

There are few online activities that need to be given new definitions. We have listed them as the follows:

An active day: A day during the course duration when online activity is undertaking, however, it has to be an active online transaction or communication, for example, sending a document, or answer a question. However, only reading the course content online is not an active day.

An active online participant/student: A participant or student has during the course accomplished at least once or more online activities, for example, sending a document or anwser a question, etc. Again, only reading the course content is not an active student.

An active activity: A participant or student has accomplished one online activity, for example, sending a message to teacher, or to other online classmates, or sending a question for online teacher. Compared with a traditional class room teaching, this action can be regarded as a comment during the lecture hour, or raising up a hand for a question, or speaking with other classmates for discussions.

A read only activity: A participant or student has during the course visited course senter once or read one page of the course's content online. This is a courseware issue: Every time when a course participant entered the course and started to read, the courseware is able to register that participant's reading activities, including number of online pages that have been browsed and time of reading. Thus, even that participant did not write any word in the course room, it is still possible to notice his/her online reading activities through our courseware. Compared with a traditional class room teaching situation, this is like a notebook for these hard

International Conference on Engineering Education   August 18–21, 2002, Manchester, U.K.

2

reading, but silent students. Consequently, read only activities could be an important part of online class performance, that supporting the course evaluation criteria.

With these new definitions for online class performance, we can look at online statistical summaries for our sampled courses.

ONLINE STATISTICAL SUMMARIES

First, we will look at NBU4 online course, conducted and finished on April 2002. Through the course database, the participants' online activities, both in writing and reading, and their final assignments were recorded. The table only displayed the data from the first two weeks of th course. As mentioned early, the course was highly practice oriented and the majority of the course content was practicing online and accomplishing assignments. One of intentions was getting participants' accomplishing required assignments and this was regarded as a part of teaching effectiveness.

TABLE I
A STATISTICAL SURVEY OF NBU4 ONLINE COURSE AT APRIL 2002

Participant	Active Activity	Read Only	Assignments
AF	2	256	1
AG	11	583	6
ED	1	45	1
AK	0	14	0
BK	5	149	2
SL	1	99	1
BN	0	56	0
RD	22	372	5
JL	4	54	0
KW	19	494	5
LF	4	400	4
EM	0	25	0
BD	5	414	2
SG	10	759	5
HG	0	30	0
TG	31	1044	5
Online teachers			
LN	3	104	
HG	35	715	
GN	24	277	

Table I illustrated a statistical survey of online course NBU4, which was conducted at April 2002. As we can see the content from the table, not only active participants and their active online activities were displayed, but also read only activities were recorded. Participant AK, BN, EM and HG had zero active activity, but from 14 to 56 frequences of read only activities, and therefore could be defined as a read only participant. Naturally, their assigment record was also zero since they never sent any document to the course senter.

As mentioned early, the successful participation for this course was directly related to the accomplishment of required assignments. The participants were required to finish 5 assignments in order to be "graduated" from the course. We can use this number as the measurement criteria.

Participant AG, RD, KW, SG and TG are these 5 participants whom did reach the goal in this course. Say,

they all passed the final exam, but how about their online class performance statistics? Table II has displayed these 5 partiicpants' online activities in frequences, both for active activities and read only.

TABLE II
A SUMMARY OF 5 SUCCESSFUL NBU4 PARTICIPANTS AT APRIL 2002

Participant	Active Activity	Read Only	Assignments
AG	11	583	6
RD	22	372	5
KW	19	494	5
SG	10	759	5
TG	31	1044	5

Looking at comparison between table I and table II, it seems to be reasonable to define a standard for this course as the following criteria:
- For active activities: If activity frequences are less than 5, failed, if more than 10, passed.
- For read only activities: If activity frequences are less than 100, failed, if more than 400, passed.

For the former, the definition is relatively obvious and easy to understand, but for the latter, the definition could be hard to understand and more diffused. As we can observe from table I, there are two other participants LF (400) and BD (414) whom did not reach 5 assignments, but both had over 400 in their read only records. In fact, the later course data indicated also participant LF indeed accomplished the 5th assignment.

To understand the above phenomenon and dilemma, it should be reminded the following facts: (a) a statistical measurement allows few exceptions; (b) what read only activities indicated was the number of web pages browsed, not necessarily the pages was read. Even it is possible to count the real number of the pages were read in syllabus, there is still no guranty for the successful accomplishment of the course, because these pages have to be understood and absorbed. Nevertheless, read only activity is still an important criterion to supporting total evaluation of a course. From the data of table I, we can surely conclude the fact that for read only activities, if activity frequences are less than 100, it is surely unsuccessful.

For further comparison, we will study another sampled online course, NBU2 conducted at April 2001. Similar as her sibling course, NBU2 database also restored also both in writing and reading information from the participants, as well as their final assignments. The course data are restored from the whole course duration, thus 3 weeks, but read only activity record was only available for the last week.

NBU2 was one of the most active online courses was even conducted throughout our university college. Majority of participants were actively engaged in the course discussions and assignments, so the final assignment number was relatively high among the participants. It will be interesting to study this case in details and see if there is any

correlations between performance and successful assignment.

TABLE III
A STATISTICAL SURVEY OF NBU2 ONLINE COURSE AT APRIL 2001

Participant	Active Activity	Read Only	Assignments
AD	7	26	6
AL	11	140	5
AN	43	196	6
AB	25	141	6
BU	24	152	8
DT	10	0	3
KT	1	0	1
MN	8	68	5
MT	16	42	7
ML	10	54	2
PT	3	0	1
RG	4	0	3
RN	27	82	7
TH	16	166	7
TN	32	89	7
Online teacher			
HG	51	57	
GN	64	84	

As table III illustrated, the majority part of pariticipants of this course did pass the course requirement for assignment accomplishment, which was 5 assignments. Of total 15 active participants, there were 10 of them did pass the course successfully, by accomplishing at least 5 assignments. Only 5 of them did not reach this goal.

At a closed look for participants, DT, KT, ML, PT and RG did not accomplished the required amount of assignments. Their active activity frequences were 10, 1, 10, 3, an 4 respectively. These numbers were obviously, below the average activity frequences of the class. Correspondingly, their read only activity frequences for the last week were 0, 0, 54, 0 and 0 respectively. For online class performance, 0 frequence means drop off from the course, which means 4 of them quiet the course at the last week.

Participant ML with 10 active frequences and 54 read only frequences in the last week, but only 2 assignments, was an exception. In fact, this participant had trouble to start with local access to internet, so this trouble delayed the whole prosess.

Since read only activity frequences was only available for the last week of the course, we need to estimate roughly a criterion for the whole course period in 3 weeks. If we consider the function of read only frequences as a linear ratio with course period, so the approximately read only frequences for 3 weeks will be 3 x frequences in last week (except these whom drop off the course).

The challge of summarizing this course, compared with previous analysis for NUB4, is great variety of active and read only activitiy frequences among the participants. It is therefore difficult to distinguish different groups, for example, successful and unsuccessful group, only by frequences since the border is not so clear. One alternative is to use average method to make distinction between groups. Say, for successful against unsuccessful groups, we can calculate the average frequences of active activity and read only for the whole group, instead of identifying the upon and lower limit of the frequences.

For NBU2 online course, if we consider the information available and combine the data from table III, it seems to be reasonable to define a standard for this course as the following criteria:

- For active activities: If activity frequences are less than 5, failed, if more than 15, passed.
- For read only activities: If activity frequences are less than 100, failed, if more than 300, passed.

Again, there were exceptional cases for these criteria, as we can easily notice from table III, AD, AL, MN for active frequences, and AD, MN, MT, RN and TN for read only frequences. These participants did not fit the above criteria though they perfectly passed the course.

One of the most important elements for standard definition or criteria establishment is empirical data, with sufficient, and only sufficient amount of data. We obviously need more empirical data in order to establish new criteria or standards.

Finally, we sampled an online course in English version, conducted at June 2001 in China. The course was enshorten into one week, and this might cause negatively for teaching effectiveness. Table IV has summarized the course activities in discussions, comments and assignments, respectively. The individual participant's online frequence data was not available for this course, so one has to study the data as a group, and only average person data is available.

TABLE IV
A STATISTICAL SURVEY OF OTIP1 ONLINE COURSE AT JUNE 2001

Participant	Discussions	Comments	Assignments
27	46	42	63
Average person	2	2	2
Online teacher			
HG	14	15	
GN	4	4	

It is easy to observe the fact that online class activity frequences are rather low for this course. With average less than 2 documents per person, the activity has to be low, even for one week.

Since the course was only conducted within one week, there was no requirement for 5 assignments accomplishment. Instead, we asked participants to fill up a certificate form in order to receive a course certificate. Finally, there were 4 participants met this requirement. Consequently, this course was not considered as successful conducting, especially from a technical aspect. However, for the introduction purpose, this was a good course for our Chinese colleagues.

To summarize the analysis of 3 sampled online courses, it has been intention to establishing and identifying a performance based evaluation standard, or criteria for online course evaluations. Two sampled online courses were used

to identifying the new criteria and the third online course was used to test the new criteria. The new criteria are merely based on online activity frequences by each course participant, in their active activity and read only frequences.

The current study only attempts to initiate a new experiment and discussion in searching for new standards and criteria for evaluations. The final establishment of such standards and criteria requires more empirical research and sufficient amount of data. However, it is possible to accumulate the empirical data from the currently sampled courses.

Based on 2 sampled online courses, which are operated in 3 week intensive interval. The online class performance evaluation can be defined as the follows:

- For active activities: If activity frequences are less than 5, failed, if more than 10-15, passed.
- For read only activities: If activity frequences are less than 100, failed, if more than 300-400, passed.

These criteria will certainly be changed whenever other course circumstances are changed, say, whenever the course extended or enshorten, or modifying assignment requirement.

Another important issue that should be discussed, is validity and reliability of the data. As the previous example indicated from read only activity frequences, we have noticed the frequences of read only were merely labeled for number of browsed web pages, not necessarily number of read pages. Criticism might question: How could you be so sure for so many pages were read, but not glanced without real focusing?

VALIDITY AND RELIABILITY ISSUES

The issues of validity and reliability are always essential for any measurement and evaluations, including online data. Previously, we have mentioned the great advantages of online data collection, compared with tradtional methods of data collection. That there is no need to design a particular questionnaire, and no need to distribute many copies of a questionnaire to respondents, and collected them back, then punching into the computer for further data analysis.

For online course, all these steps were not necessary since the data was restored directly in the course's database, and respondents do not have to pretend to give the right answers, so the answers do not bring his/her real meaning. In a aspect, online data collection method provides better validity and more reliability for the data analysis.

On the other hand, there are few weak points for these direct online data, which are listed as the follows:

- First, they are primary secondary data, so they could not serve for the certain purpose. For some research work, it is still necessary to design questions and construct the questionnaire in order to get the correct answers.

- These data are quantitive and statistical based number games, so it requires large population and samples to be valuable.
- The frequences of online activity are based on number of documents, not content of each document. This will make huge differences between individuals, say, a person who write many short messages would get better frequences than a person who prefer to write a long, but desent article.
- There is an opening for overlapping document, so that the participants whom are not familar with courseware might be able to submit duplex, triplex or even multiplex copies of one single document. Figure 1 illustrated a such example. In reality, there is only one "In Porgress YNW......" and one "In Progress LZH....." needed. Consequently, such overlapping document results incorrect calculation of statistical survey and error in data quantity.

Module 3-What so Special for Online Classroom Exercise (due: 21.06.2001)
Submitted 22.06.2001 LTN on 22.06.2001
Submitted 22.06.2001 LNG on 22.06.2001
In Progress TYM on 22.06.2001 <Public and Finance>
In Progress JLP on 21.06.2001 <Education and Economics>
In Progress YNW on 21.06.2001
In Progress YNW on 21.06.2001
In Progress YNW on 21.06.2001
In Progress YNW on 21.06.2001
In Progress WCG on 21.06.2001<Information and Resources and Foreign>
Submitted 21.06.2001 DXG on 21.06.2001
In Progress LZH on 20.06.2001
In Progress LZH on 20.06.2001
In Progress LZH on 20.06.2001
In Progress LZH on 20.06.2001
In Progress DNG on 20.06.2001
In Progress ZHP on 20.06.2001

FIGURE. 1
AN EXAMPLE OF OVERLAPPING DOCUMENTS IN ONLINE COURSE.

People may wonder: Traditional designed questionnaire versus online direct data collection, which one porvides better validity and reliability data collection and analysis? Which one should be applied in our teaching effectiveness evaluations?

Like other academic issuesand debates, there is no fixed answer for one or the other. It depends perhaps on what you wish to analyze, or which particular circumstance of the course belongs to.

Generally, if an online course has sufficient amount of data and many participants, it will be too costly and too slow to design a standard questionnaire for data collection. At the same time, large amount data provides better validity and reliability for online data collection, so it will be easy, quick, but also reliable to collect data direct from the course database.

Online teaching is a new approach, though unfamiliar for many. Still, there are a lot of openings and possbilities to

combine with traditional data collection. For example, it will be much easy to design a questionnaire and paste it directly into the online courseware, so that the students would be able to respond the questions online. In many cases, it will probably be easier to construct a questionnaire directly in the courseware, than on the paper.

It is therefore reasonable to believe that validity and reliability issues will be be solved gradually.

COULD A NUMBER GAME REPLACE THE EXAM?

In this article, we have made effort to identify, define, experiment and discuss about a quantity based standard or criteria for online class performance evaluation. The intention is, not yet, to find such a standard or criteria, but to experiment and discuss the issue.

However, based on our 2 sampled online courses, both with 3 week intensive interval. The current study has identified the following class performance standards or criteria:

- For active activities: If activity frequences are less than 5, failed, if more than 10-15, passed.
- For read only activities: If activity frequences are less than 100, failed, if more than 300-400, passed.

Naturally, the validity and reliability of the above standards or criteria are debatable. Nevertheless, it is always interesting to question about the future: Could a number game like this replace the final exam for our future courses?

Again, it will be hard to provide a fixed answer for this question. At the present time or near future, it will be hard to think about the replacement possibility. The reasons as we discussed in the article, are because of current validty and reliability of online data collection are still in pre-matured stage, so there are many questions and issues need to be answered and solved.

A good answer for this question may be that a number game like this could support the final exam for our courses.

By using a number game like this, it will be easy to extract or track up the wanted candidates for final exam acception or rejection.

By using a number game like this, it will be possible to combine with the final exam results and provide each student a total mark that may illustrate a more correct profile of that student.

By using a number game like this, it will be convenient to collect data for conducting research work or comparing with other courses.

There are a couple of weak points of this number game, that were mentioned in the validity and reliability debate. They could actually be rewrite as further suggestions for the improvement or future work:

- There are many openings and possibilities in the courseware for creating of questionnare. In case of using questionnaire, rather using through the online course, thus, courseware. This will avoid puching and make data collection much easier.
- Make sure there will be sufficient number of participants, say, minimum 15 persons, and activity frequences, say, at least 5 questions and comments are required from each person in each of 10 chapters. Thus, there will be 15x5x10=750 frequences as the basis.
- Be aware of individuals style of writing, say, many short messages versus a long, but desent article. The former may get better frequence account than the latter. But the latter would be leading in the content. Remember to balanse them.
- Eliminating overlapping document quickly, so that the statistical calculation of survey is becoming correct. The participants and students should also receive good training in use courseware,so that they are really familiar with courseware.

The number's consequences are obvious for online courses. With helps of number, it will be easy to extract or identify the correct profile of an online course. Though it will be too early, or never actual to replace the final exam with a data based evaluation, there are still many openings and usages for online class performance statistics, and the numbers are the essential sources to this evaluation.

International Conference on Engineering Education　　　　　　August 18–21, 2002, Manchester, U.K.

6