AC 2010-1321: EVALUATION OF PEERWISE AS AN EDUCATIONAL TOOL FOR BIOENGINEERS

Paul Denny, University of Auckland

Paul Denny is an instructor in the Computer Science Department at the University of Auckland in New Zealand. In addition to Computer Science Education, his research interests include collaborative student learning and he created the PeerWise tool to support this approach.

Beth Simon, University of California, San Diego

Beth Simon is a Lecturer with the Potential for Security of Employment in the Computer Science and Engineering Department at the University of California, San Diego. Her research interests include educational technology and computer science education research.

Melissa Micou, University of California, San Diego

Melissa Micou is a Lecturer with the Potential for Security of Employment in the Department of Bioengineering at the University of California, San Diego. Her interests include identifying novel strategies to integrate teaching and research and enhancing the recruitment and retention of underrepresented populations in engineering. Dr. Micou is the program director of an NSF-sponsored Research Experience for Undergraduates (REU) program in Regenerative Medicine, Multi-Scale Bioengineering, and Systems Biology.

Evaluation of PeerWise as an Educational Tool for Bioengineers

Abstract

There is a need to develop, validate, and widely implement tools that incorporate proven educational strategies including collaborative learning, active learning, and peer tutoring. PeerWise is an innovative, web-based system in which students create multiple-choice questions and answer those created by their classmates. Creating the question bank, rather than just accessing an existing one transforms students from passive recipients to active learners. The objectives of this study were to characterize the use of PeerWise in a lower division bioengineering course and to measure its efficacy in improving student learning. Students who used PeerWise performed significantly better on the final exam compared to minimally active users. PeerWise was easy to implement, placing almost no burden on the instructor, and the majority of students in the class used it voluntarily to study for their exam. These findings demonstrate the potential of PeerWise to be a powerful and widely used educational tool for bioengineers.

Introduction

Research in science and engineering education has identified strategies to improve student learning, including: learning communities, collaborative learning, active learning, peer review, and peer tutoring[1-4]. There is a need to build on this knowledge by developing, validating, and widely implementing educational tools that incorporate these proven strategies.

PeerWise is an innovative, web-based tool that exploits the familiarity students have with user-generated content, as exemplified by popular sites such as YouTube, to create an online learning community. Specifically, PeerWise provides a framework for students to work collaboratively with their classmates to create a large repository of multiple-choice questions. Creating the question bank, rather than just accessing an existing one, transforms students from passive recipients to active learners. To author a high quality question a student must understand the topic in depth, consider misconceptions when proposing distracters (plausible but incorrect answers), and write a clear explanation for the correct answer. Once a question has been submitted, other students in the course can use the question for self-assessment, learn from the explanations, evaluate the quality of the question, and participate in a discussion thread by offering additional explanations or posting comments. The process of evaluating the quality of a question and corresponding explanation require behaviors at the very highest level in Bloom's taxonomy of the cognitive domain [5,6]. By establishing a learning community that incorporates collaborative learning, active learning, peer review, and peer tutoring, PeerWise has the potential to be a powerful teaching tool.

Since its development in 2007 at the University of Auckland, more than fifteen thousand students at twenty-three different institutions have used PeerWise worldwide. Rigorous evaluation of PeerWise in computer science courses has shown that student-generated questions cover the full range of topics within a course and are generally high quality [7,8]. Student attitudes toward PeerWise are overwhelmingly positive as demonstrated by the finding that students answer many more questions than their instructor requires them to and voluntarily use the system in preparation for exams [5,9]. Computer science students who actively use PeerWise have also been shown to perform better on final exams than students who are not active [5].

Assessment of PeerWise in the field of computer science has demonstrated its potential as a teaching tool. However, the nature of computer science inherently favors logical thinking, computer science content fits nicely in a multiple-choice format, and computer science students will readily use computer-based tools. Therefore, use of PeerWise in other fields does not guarantee success. Inherent differences between computer science and other engineering and science disciplines necessitate further investigation of PeerWise.

Evaluation of PeerWise as an educational tool for bioengineers and biomedical engineers (BE/BMEs) was initiated in a biostatistics course taught within the Bioengineering Department at the University of California, San Diego with the following rationale. Results from a comprehensive survey conducted by the Vanderbilt, Northwestern, Texas, and Harvard/MIT Engineering Research Center (VaNTH-ERC) show that BE/BME experts both from academia and industry identified "descriptive statistics" and "hypothesis testing" among the top three most important concepts (out of 274) within an undergraduate BE/BME curriculum [10]. The same survey showed that statistics is one of the most commonly required courses for undergraduate BE/BMEs, with 78% and 81% of accredited and non-accredited programs requiring a statistics course, respectively [10]. Finally, feedback at our own institution indicates that students have difficulty in biostatistics because the problems available in typical textbooks emphasize calculations rather than concepts; students want different types of practice problems than what is readily available. The multiple-choice question format used by PeerWise works well for that purpose. Therefore, the objectives of this study were to characterize the use of PeerWise in an introductory biostatistics course for bioengineers and to measure its efficacy in improving student learning.

Methods

Description of the System

PeerWise is a web-based repository of multiple-choice questions specific to a particular course with alternatives and explanations written by students. After logging in, the main menu shows three subsections entitled: "Your Questions", "Unanswered Questions" and "Answered Questions" (Figure 1A).

"Your Questions" allows a student user to review all the questions they have authored. When contributing a question, the author must provide a question stem, two to five alternative answers, an indication of the correct answer, and an explanation. As soon as a question is contributed it is available in the "Unanswered Questions" section for other users enrolled in the course.

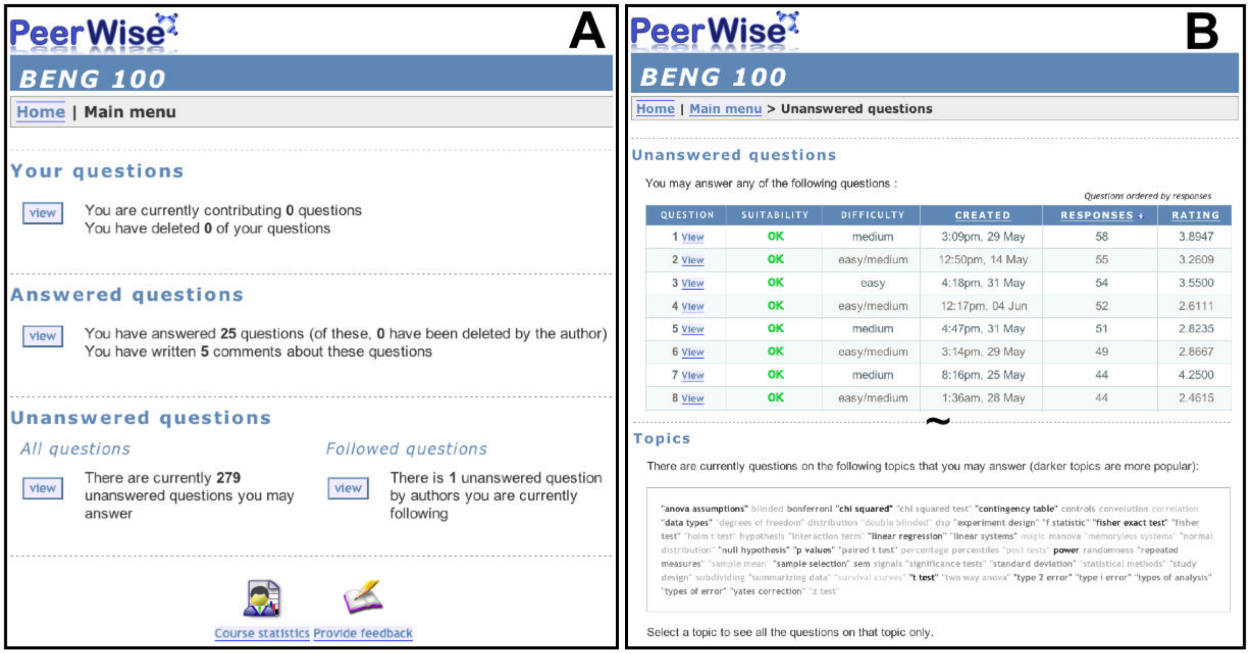

Figure 1: Examples of the PeerWise interface. A) The main menu is divided into 3 sections: the questions that the student has contributed, the questions contributed by others that the student has answered, and the questions contributed by others that the student has not answered. B) A list of all the questions the student has not answered is provided and can be sorted by multiple features.

"Unanswered Questions" are presented in a table (Figure 1B) that shows their suitability, level of difficulty, date created, number of responses from other users, and quality rating. The level of difficulty and quality rating are based on user feedback. Suitability is determined by comparing the author's suggested answer with the most popular user answer, also taking into account the current quality rating. The table can be sorted by any of the last 3 columns. A list of topics linking to unanswered questions on that topic is at the bottom of the page. The topic list is generated automatically by PeerWise based on the content of the question bank. The user can select which question to answer based on level of difficulty, quality rating, number of responses, topic, or another criteria. When a user selects an unanswered question they are shown the question stem and possible answers. After the user selects the answer they believe is correct they are shown the author's suggested answer, a histogram of all responses, the author's explanation, and comments submitted by other respondents. The user then has an opportunity to categorize the question difficulty as "easy", "medium", or "hard", rate the quality of the question on a scale from 0 to 5, and provide a comment. Answered questions are permanently available for future review in the "Answered Questions" section.

"Answered Questions" are presented in a table similar to those in the "Unanswered Questions" section except that the suitability column is replaced with a results column. This column shows whether the user answered the question in agreement with: both the author and majority of respondents, the author, the majority, or neither. As more students provide responses the accuracy of the results column improves.

Two additional features of the system are a leaderboard and the ability to "follow" questions submitted by a particular, yet anonymous, author. The leaderboard displays the top rated questions, the rating of questions submitted by the user, the most questions answered by a student, and other statistics. A student may elect to "follow" a specific author if they find the author's questions to be particularly useful. Questions submitted by the followed author are listed separately from other unanswered questions.

Evaluation of PeerWise

Efficacy of PeerWise was investigated in a second-year bioengineering course required for all bioengineering majors except those specializing in bioinformatics. A midterm exam including multiple-choice, short answer, and free response questions was administered approximately halfway through the quarter. Although the course covers a variety of topics, all the lectures before the midterm are devoted to biostatistics and experimental design. *After* the midterm exam, PeerWise was introduced and students were required to author a minimum of two statistics questions as a small component of their course grade (<1%). Students answered PeerWise questions, rated question quality, and provided comments on a voluntary basis since these activities were not factored into their grade. At the end of the quarter, a comprehensive final exam, also consisting of multiple-choice, short answer, and free response questions, was given.

Student Learning

Exam scores, a universal mode of assessment in undergraduate courses, were used to measure student learning. Raw scores on the midterm and final exam were linearly transformed into normally distributed T-scores with a mean of 50 and a standard deviation of 10 [11]. This transform puts scores from different exams on the same scale, allowing them to be compared regardless of differences in the mean or variability of the raw scores[11]. Data from students who took both the midterm and final exams in spring 2009 were used in this analysis (152 out of 156 enrolled students). T-scores from the midterm and final exams were used for statistical analysis as described below.

Student Activity

Students created a PeerWise account so the number of questions authored and answered could be automatically logged for each student by the system. The number of questions answered by the end of the quarter was used to classify a group of highly active (HA) and minimally active (MA) users. Exactly 50 students answered no PeerWise questions and these students constituted the MA group. The 50 students who answered the most questions were defined as the HA group. Both groups were roughly one-third the size of the entire class.

Statistical Analysis

To test the hypothesis that PeerWise improves student learning, a mixed model repeated-measures ANOVA was performed (α=0.05) using GraphPad Prism Software. The two main factors considered were PeerWise activity level (HA or MA) and exam (midterm or final). Exam was analyzed as a repeated-measure so that the midterm and final scores of an individual student were paired in the analysis. Post-hoc analysis was performed using Holm's t-tests (α=0.05). Data is reported as mean±standard deviation or median depending on whether the data is normally distributed.

Results

Student Use of PeerWise

Nearly all students completed the requirement of authoring a minimum of 2 PeerWise questions resulting in a repository of 327 different questions. Among the most common question topics identified were: *types of data, statistical power, null hypothesis, p-values, t-tests, F-statistic, analysis of variance assumptions, $\chi 2$ / contingency tables, Fisher exact, and linear regression.* Sixty-seven percent of students used PeerWise voluntarily as a study aid by answering questions from the question bank. The total number of responses submitted by these students was 4,154. The median number of questions answered per student was 22 and ranged from 1 to all of the questions. These students commented 197 times on questions submitted by other users.

The three examples below illustrate the type and varying quality of student-authored questions.

Example 1 was answered by 57 users, categorized as "medium/hard", and had the third highest quality rating of the questions submitted (4.4 out of 5.0) based on 22 student ratings. Two students provided comments on the question. The most popular answer, selected by 39% of the respondents, was the best answer.

> Question: *For a given test, if you keep the sample size the same but make α smaller (require stronger evidence before reporting that a treatment has an effect), this has the effect of:*
> *A increasing the power of the test*
> *B decreasing the likelihood of making a Type II error*
> ***C decreasing the power of the test***
> *D increasing the likelihood of making a Type I error*
> *E Both A and B*
>
> Author Explanation: *Type I and Type II errors are intertwined. As you require stronger evidence before reporting that a treatment has an effect, i.e., make α smaller, you also increase the chance of missing a true effect, i.e., make β bigger or power smaller. The only way to reduce both α and β simultaneously is to increase the sample size, because with a larger sample size you can be more confident in your decision, whatever it is. Glantz, p.186* [12].
>
> Student Comments*: 1) Good question. 2) Good conceptual question and thank you for the reference page.*

Example 2 was answered by 45 users, categorized as "easy/medium", and had the median quality rating (3.0 out of 5.0) based on 17 student ratings. Two students provided comments on the question. The most popular answer, selected by 62% of the respondents, was the best answer.

Question: *Which of the following cases does not cause the t test to be used incorrectly?*
A The means of three or more groups are to be compared.
B The dependent variable is not continuous.
C The samples contain the same individuals.
D There are two or more independent variables.
E The samples do not have the same sample size.

Author Explanation: *A) is wrong because the t-test only compares the means of two groups. B) is incorrect because the dependent variable must be continuous. C) is wrong because the samples or groups cannot contain matched subjects. D) is incorrect because there must only be one independent variable. E) is correct. Samples with different sample sizes can still be compared by the t-test as there are separate values of n in the t-test equation for the sample sizes of both samples.*

Student Comments: *1) confusingly worded. 2) Good question, but didn't specify paired or unpaired t-test, in which case C can be a right answer (E is still the best answer).*

Example 3 was answered by 9 users, categorized as "medium", and had the lowest quality rating possible (0.0 out of 5.0) based on 2 student ratings. One student provided a comment on the question. The most popular answer, selected by 44% of the respondents, was different than the answer provided by the author. Based on this inconsistency the question was flagged as potentially unsuitable. There is no clearly correct answer.

Question*: A study is done to see if elbow injuries are more common in sports where rackets are used. In this case, the sports under study were tennis and racquetball. The control is a sport that does not involve the use of rackets, specifically basketball.*

Group	***Injury***	***No Injury***	***Total***
Control (basketball)	*9*	*40*	*49*
Tennis	*35*	*12*	*47*
Racquetball	*47*	*15*	*62*
Total	*91*	*67*	*158*

Which of the following tests/procedures would allow you to make the most accurate and specific conclusions?
A Chi Squared Test
B Yates Correction
C Subdividing Contingency Tables
D A and B
E All of the above

Author Explanation: *A) A chi squared test alone will only tell you the probability of a significant difference between the groups, not specifying which group(s) is(are) different. B) A Yates correction is a fix for chi squared test when you deal with 2 x 2 table (when degrees of freedom equals 1). If a Yates correction is done, it means that a chi squared test is done, therefore Yates correction alone is incorrect. C) A subdividing contigency table will allow you to make the most accurate and specific conclusions. By splitting the* *data into two sets of 2x2 contigency tables, you will be able to determine which specific groups are significantly different. However, a subdividing contigency table involves the use of the chi squared test and Yates correction, therefore...E) All of the above is the correct answer.*

Student Comment: *I disagree with your answer. You must consider your control as part of the contingency table, thus you have 2x3 contingency table and wouldn't need a Yates correction.*

<u>Characterization of the HA and MA groups</u>

As described previously, the MA group was composed of students who did not answer any PeerWise questions. In contrast, the median number of PeerWise questions answered between the midterm and final exams in the HA group was 52.

<u>Student Learning</u>

Exam scores were dependent on the interaction between PeerWise activity level and exam ($p<0.05$) but not on activity level or exam alone ($p=0.15$ and $p=0.71$, respectively). Post-hoc analysis showed that students classified as HA and MA performed similarly on the midterm exam, which was administered before PeerWise was introduced. Students who used PeerWise between the midterm and final exams performed significantly better on the final exam (Figure 2, $p<0.05$). Specifically, students in the HA group *increased* their exam T-score between the midterm and final by 1.6±7.3 points while T-scores in the MA group *decreased* by 2.1±7.9 points. The difference between the average T-scores of HA and MA students was 3.7 points, which corresponded to 4.5% on the exam raw scale.

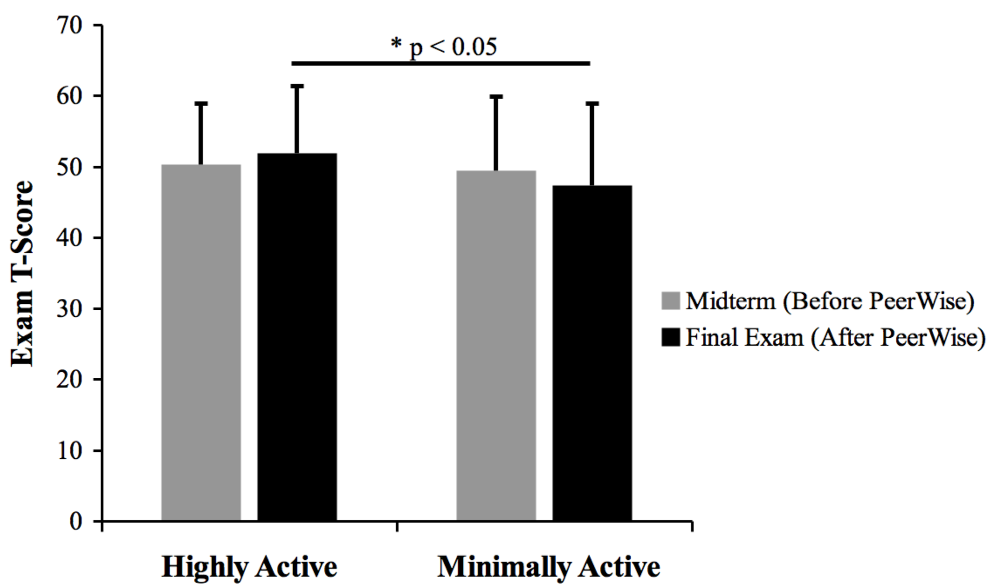

Figure 2: Exam T-scores from the midterm and final exams for highly active (HA) and minimally active students (MA). Raw exam scores were transformed to generate a normal distribution and to allow comparison of scores regardless of differences in the mean or variability of the raw scores on the two exams[11]. Data are shown as mean±SD with n=50.

Discussion

Previous studies to measure the efficacy of PeerWise in improving student comprehension have been limited and only in the field of computer science [5,7-9]. Here we demonstrated that students in a biostatistics course who used PeerWise between the midterm and final exams performed significantly better on the final exam than students who did not. This research is an important first step toward demonstrating the effectiveness of PeerWise as an instructional tool for BE/BMEs.

The traditional use of multiple-choice questions has been criticized as encouraging surface learning and favoring the development of skills at the lower end of Bloom's taxonomy [6,13]. It follows that, for PeerWise to be effective at achieving higher level learning objectives, the questions should not be simple recall, relying on only the memorization of facts. When assigning students to contribute two questions to PeerWise, the instructor only *briefly* mentioned that the questions should test more than simple recall; nevertheless, the complexity of the resulting questions and explanations was impressive. The finding that active users of PeerWise performed better on a comprehensive final (where free response and short answer questions accounted for 45% of the points and high-level, conceptual multiple-choice questions primarily accounted for the rest) suggests that active users did develop a deeper understanding of the course material.

Any successful teaching tool will function well in a variety of courses and will be easy for instructors to adopt and use. PeerWise is online and fully automated making it as easy to use in large and small courses alike. Commonly offered BE/BME courses that readily lend themselves to the PeerWise multiple-choice format include: Biomaterials, Physiology, Computing, Tissue Engineering, and most basic science courses. In more quantitative courses, such as Biomechanics, Bioinstrumentation, and Heat & Mass Transfer, PeerWise can be used to emphasize concepts rather than calculations. Adding this dimension to a course may help move students away from the common "plug and chug" mentality. Instructors who wish to use PeerWise only need to request an account online, forward registration information to their students, and perhaps provide a small grade incentive. The ability of students to rate the question quality and to select unanswered questions based on these ratings leads to better questions being answered much more frequently, a trend also observed here [7]. As such, additional monitoring of the questions by the instructor is not intended or necessary.

Use of PeerWise in a course provides numerous benefits for the instructor. Unlike the traditional engineering disciplines, there is a lack of commercially available teaching material specifically for BE/BME. Instructors routinely use traditional engineering textbooks and supplement them with their own examples of BE/BME applications. PeerWise is an efficient way of generating a large amount of material that could be useful to the instructor for preparing future homework, exams, in-class examples, or published course materials. Instructors may wish to monitor the PeerWise content to gain valuable feedback regarding specific student misconceptions and to identify general topics that students do not understand.

While a demonstrated improvement in test scores is an encouraging finding, further study may be necessary to convince a large number of instructors to adopt the technology. Toward this end, it is important for experts to evaluate the quality of the student-generated question bank, particularly with respect to achieving higher-level educational outcomes. Verifying the observation that students chose to answer high quality questions more frequently than the low quality questions will further reassure educators that the tool is being used productively [7].

It has been suggested that HA students spend more time studying in general, are inherently better students, and have a greater capacity to earn high scores. However, the finding that students in the MA and HA groups scored similarly on the midterm exam suggests that the academic ability of students in two groups are generally similar. It is possible that, after the introduction of PeerWise, HA students spent more total time studying than their MA counterparts. If that is the case, then PeerWise may be effective, in part, because it provides students with additional study materials in a format that they prefer to those traditionally available. In fact, sixty-seven percent of students in this study used PeerWise for activities that were not a required component of their course grade. This result supports an earlier study showing that students in an introductory computer science course answered many more questions than they were required to and voluntarily used the system to prepare for exams [9]. Additional qualitative data is needed to fully characterize student attitudes about PeerWise and to determine if HA students devoted more total time to studying for the final than MA students.

The data for this study was collected from all students enrolled in the course and there were many uncontrolled variables, as is reflected in the high standard deviation of the data. As such, the experimental samples represented the diversity of a typical student population including different learning styles and levels of academic achievement. While PeerWise was shown to be effective for the HA group as a whole, it would be helpful to characterize which types of students tend to utilize and benefit more from PeerWise.

In summary, PeerWise improved the performance of BE/BME students on a comprehensive exam. Its use requires minimal effort and it is helpful for the instructor, making it an attractive teaching tool for a wide variety of disciplines and educational environments.

References

1. Brower AM, Dettinger, Karen M. What IS a Learning Community? About Campus. 1998(November-December):15-21.
2. Courter SS, Johnson G. Building Community and Retention Among First-Year Students: Engineering First-Year Interest Groups (eFIGs). 37th ASEE/IEEE Fontiers in Engineering Education. Milwaukee, WI, 2007.
3. Dochy F, Segers M, Sluijsmans D. The use of self-, peer and co-assessment in higher education: A review. Studies in Higher Education. 1999;24(3):331 - 50.
4. Prince M. Does Active Learning Work? A Review of the Research. Journal of Engineering Education. 2004;93(3):223-31.
5. Denny P, Luxton-Reilly A, Hamer J, Purchase H. PeerWise: Students sharing their multiple-choice questions. International Workshop on Computing Education Research. Sydney, Australia: ACM, 2008.
6. Krathwohl DR. A revision of Bloom's taxonomy: An overview. Theory into Practice. 2002 FAL;41(4):212-8.
7. Denny P, Luxton-Reilly A, Simon B. Quality of student contributed questions using PeerWise. In: Hamilton M, Clears T (eds). Eleventh Australasian Computing Education Conference. Wellington, New Zealand: CRPIT, 2009.

8. Denny P, Luxton-Reilly A, Hamer J, Purchase H. Coverage of course topics in a student generated MCQ repository. Innovation and Technology in Computer Science Education. Paris, France: ACM, 2009.
9. Denny P, Luxton-Reilly A, Hamer J. The PeerWise system of student contributed assessment questions. In: Hamilton M, Hamilton S (eds). Tenth Australasian Computing Education Conference. Wollongong, Australia: CRPIT, 2008.
10. Gatchell D, Linsenmeier R. VANTH Biomedical Engineering Key Content Survey, Part Two. American Society of Engineering Education. Honolulu, HI, 2007.
11. Francis RW. Common Errors in Calculating Final Grades. Thought & Action. 2006(Fall):15-22.
12. Glantz SA. Primer of Biostatistics. 6th ed. New York: McGraw-Hill Companies Inc.; 2005.
13. Scouller K. The influence of assessment method on students' learning approaches: Multiple-choice question examination versus assignment essay. Higher Education. 1998 JUN; 35(4):453-72.