Pedagogical Implications of Score Distribution Pattern and Learner Satisfaction in an Intensive TOEIC Course

Che Chang Kang[1]

[1] Department of English, National Changhua University of Education, Changhua, Taiwan

Correspondence: Che Chang Kang, 7F., No. 80, Taihe Central St., Changhua City, Changhua County 500, Taiwan. Tel: 886-975-608-991. E-mail: paxsonkang@gmail.com

Abstract

The study aimed at investigating TOEIC score distribution patterns and learner satisfaction in an intensive TOEIC course and drew implications for pedagogical practice. A one-group pre-test post-test experiment and a survey on learner satisfaction were conducted on Taiwanese college EFL students (n = 50) in a case study. Results showed that the intensive TOEIC course statistically improved students' total scores and listening scores, while its effectiveness in improving reading scores was statistically insignificant. The descriptive score distribution pattern further illuminated the issue of outliers, e.g. those who progress or regress significantly as compared to the whole class. Moreover, the learner satisfaction analysis proposed a multiple regression model that indicated three positive predicting factors for students' satisfaction on the course: articulation, textbook and facility. The findings provided implications in two aspects. First, a more eclectic, individualized teaching approach should be taken to meet not only average but also outlier students' needs in intensive TOEIC courses. Secondly, course designers are provided with concrete directions for improving students' satisfaction on the course, which entails potential values in increasing learning motivation.

Keywords: TOEIC course, score patterns, learner satisfaction

1. Introduction

1.1 Standardized TOEICs versus Non-Standardized Learners' Variables

Objective evaluation of language proficiency depends on a good testing mechanism. Theoretically, Brown's (2000) assertion best facilitated our understanding of a good, affordable test. As suggested by Brown (2000), three classic criteria for a "good" test are practicality, reliability and validity. In real practice, using a "good" test to evaluate individuals' proficiency can be challenging considering the fast-changing and diversified components of learners' variables (Richards & Rodgers, 1986; Brown, 2000; Harmer, 2007). With this dynamic background, we have seen increasingly growing popularity of English certificate courses used by schools and institutes in Taiwan and abroad (Phy, 2010). Among these certificate tests, the TOEIC stands out for living up to the afore-mentioned theoretical traits of a "good" test (Sewell, 2005; Phy, 2010; Jee & Kim, 2013). According to the data provided by the official website (Note 1) of Educational Testing Service (ETS) in Taiwan, in 2012 around the world seven million people took the TOEIC, with more than 14, 000 companies, schools and government organizations adopting the TOEIC as assessment tools in 150 countries. A Taiwan-based statistics (Note 2) also demonstrated a steady growth in the number of test-takers from 68,054 in 2005, 141,740 in 2007, and 162,968 in 2009, to the more recent 237,322 in 2011. Under this global trend, TOEIC courses aiming at helping students get higher scores have mushroomed quickly in both Taiwan and most Asian countries in recent years (Jee & Kim, 2013). Nevertheless, the dominant use of standardized TOEICs as language proficiency assessment and the use of TOEIC courses for steering language learning in EFL context need to be treated carefully considering the multi-faceted nature of learners' variables, like attitudes, interests, motivation, learning styles and so forth (Savignon, 2002).

In one sense, the growing trend of using standardized TOEICs to assess learners' overall language proficiency deserves further contemplation. In another, we are supposed to exert equal amount of caution in designing an effective TOEIC course that suits the needs of learners and prepare them for taking TOEICs. Specifically, it's important that we take advantage of our understanding of learners' variables, self-reported feedback, needs, satisfaction, etc. to examine the effectiveness of a TOEIC course and to, ultimately, improve its quality. To this

end, a brief look at learner-centered studies in this regard is needed. By referring to a methodological history of language teaching that emphasized learners' variables, we have been taught to be cautiously eclectic in making enlightened choices in the best of what we knew and what we can teach (Richards & Rodgers, 1986; Larsen-Freeman, 1986; Brown, 2000). As Harmer (2007) proposed, today's teaching can be a process of choosing between the best elements of a number of different ideas and methods. Under more scrutiny, an eclectic combination of teaching methods, nevertheless, doesn't pick a quick equivalent to being a disorganized ragbag of different activities with no obvious coherence or philosophy to underpin them (Harmer, 2007). Instead, the overt complexity of learners' factors should be more suitably addressed with a principled eclecticism that caters to needs emerging from a wider set of individual factors (Savignon, 2002; Harmer, 2007).

The discussion in the current study is presented against this background and the conceptual framework is underpinned by previous studies that emphasized learners' variables as a core subject in language teaching and learning (Richards & Rodgers, 1986; Brown, 2000; Harmer, 2007; Savignon, 2002). Under the broad spectrum of learners' variables, the current study focused on learners' satisfaction, which was implied as indispensible in the broader need-assessment survey of learners (Lin, 2012; Savigon, 2002). The notion of course satisfaction in the current study was formulated upon the Learner Attitude toward Selected Course Components (LASCC) proposed by Savignon (2000, p. 127). Based on Savignon's (2000) theoretical framework, the construct of satisfaction is defined in a broader sense, composed of a host of factors that are related to the essentially complex nature of EFL learning environment, both linguistically and non-linguistically (Robb & Ercanbrack, 1999; Lai, 2008). As argued by Savignon (2000), needs-assessment survey has been used successfully as a first step in the revision of existing programs. By establishing a slightly different set of questionnaire items that suit the context of TOEIC learning, the current study represented a small effort in investigating ways of improving the quality of TOEIC courses from a learner satisfaction point of view, which further hold implications in pedagogical realm.

2. Literature Review

2.1 The Use of TOEIC in Asia: Effect and Washback

In recent years, universities in Taiwan have widely adopted TOEIC, one of the many designated language tests, as the language proficiency "graduation threshold". The fact that numerous companies and institutes have extensively adopted TOEIC scores for recruitment might also explains the growing size of intensive TOEIC courses across college levels of education (Phy, 2010). The popular use of TOEIC as evaluation standard is not restricted in Taiwan. Korean companies are also among those that use it for hiring new staff (Phy, 2010). In a study examining the reliability and validity of the TOEIC in Korean contexts, it was proposed that the TOEIC seems to be a relatively reliable and valid test of listening and reading comprehension for untrained first time test takers (Sewell, 2005). The rather friendly written testing items, grammar, structure, reading and writing (selective) might be the reason why it becomes welcome among examinees (Phy, 2010). A powerful relationship has also been found between language testing, teaching and learning (as cited in Booth, 2012). Considering the fast growing of the TOEIC, the TOEIC courses, and, more importantly, its well-recognized acceptance as a reliable evaluation tool, it's important to conduct more in-depth research on its potential value as well as caution that instructors and students should be aware of.

While the practical benefits of taking TOEIC as a job-seeking booster are known to both the examinees and those who promote such testing, there is still concern about its washback (Booth, 2012). Washback, a term popular in British applied linguistics and commonly referred to as backwash in the field of general education, may be understood as the influence that a test has on teaching and learning (Robb & Ercanbrack, 1999). In a recent study, Booth (2012) evaluated students' perception on TOEICs through focus group methodology in a Korean university setting. His efforts in using the Activity Theory in exploring what's behind the TOEIC course highlighted the significance of situated learner goals and actions, and the influence of learner histories, cultural and social associations, attributes, conceptualizations, and agency in directing the washback effect of a test.

2.2 The Effectiveness of TOEIC Courses

Studies regarding the effectiveness of TOEIC preparation on students' improvement in scores were bleakly sparse (Robb & Ercanbrack, 1999). In a study examining the effect of direct test preparation on the TOEIC scores of Japanese university students, it was found that the improvements that students make in listening and reading parts of TOEIC are different, with English majors making equal improvements in listening and reading parts while non-majors progress more in reading than in listening (Robb & Ercanbrack, 1999). The unstable correlation between improvement in scores and the background of test-takers were attributed to numerous reasons, including outside-the-classroom preparation, the instructors' talk, technical issues regarding the administration of the test, etc.(Robb & Ercanbrack, 1999).

The non-linear relationship was also found between scores and real language competence, as examined by Cunningham (2002). Results suggested that a correlation neither exists between TOEIC test scores and communicative abilities, nor between TOEIC test score gains and improved communicative competence. It was also suggested that TOEIC test-preparation does not result in more accurate use of structure (Cunningham, 2002).

While the washback incurred in using standardized tests for language learning is controversial, i.e. negative or positive (Robb & Ercanbrack, 1999), the recurring divergence has contributed to studies addressing a more feasible way to combine authentic materials with test teaching. Phillips (2006) sought solutions for university English language instructors who wanted to aid students in improving their TOEIC listening scores, while maintaining a focus on the use of English in natural contexts in their listening and speaking classes. In line with this discussion of teaching TOEIC from a learner's perspective, Miyamoto (2010) conducted similar research on a TOEIC preparation course in a Japanese college. As Miyamoto (2010) stated:

The preparation courses for the TOEIC® test are conducted in such a way that an instructor gets the students to work on sample questions on preparation texts. However, this instruction style and method do not necessarily meet the needs of the students. (p. 71)

Overall, teaching of TOEIC and the implementation of TOEIC courses has received a modified definition. The trend toward exploring a different role of TOEIC/TOEIC course in today's multi-dimensional learning environment becomes more obvious than before. However, the studies were mostly conducted in Korean or Japanese contexts. Studies aimed at exploring the implementation of TOEIC course in Taiwanese context were limited. One of the few studies done in Taiwanese context was conducted by Lai (2008), who examined the effectiveness of college English-featured course on TOEIC. The major findings were closely connected to the previous line of discussion regarding learners' variables in TOEIC teaching. It was found that key reasons why the direct test preparation failed to significantly increase the post-test score gains included test-taker factors, nature of text, speaker variables, socio-cultural competence and strategic competence (Lai, 2008).

Under the shift of trend from a teacher-centered to a student-centered approach, the previous studies were central to presenting an overview of the effectiveness, implementation difficulty, and the association of students' multiple factors with TOEIC teaching. However, in the review of literature, it deserves some note that the analysis of scores only focused on its inferential statistical results, which doesn't provide rich insights about individual language learners (Johnson, 1992). It's hypothesized that the descriptive analysis of the score pattern distribution could better reveal the outliers cases in terms of students' score gains. In addition, among previous studies that examined learner variables, there was no specific focus on one particular construct. Identifying these shortcomings, the current study seeks to investigate the effectiveness of TOEIC course through a different approach that focuses on students' score pattern distribution, both inferentially and descriptively, and the construct of course satisfaction in a case study in Taiwanese EFL context. A different approach used in the current case study might serve to answer the following research questions:

1) Can the intensive TOEIC course improve students' scores? What are the score patterns before and after taking the course?

2) What are the factors that are associated with or serve to predict students' overall satisfaction with the course?

3. Methodology

3.1 Procedure

The study is a one group pretest-posttest case study, comprising four stages. Firstly, the researcher took advantage of the implementation of an intensive TOEIC course as the body of the experiment. Secondly, the participants enrolled in the course took a pre-test before the course started. Thirdly, the intensive TOEIC course began and lasted for one month. Fourthly, the participants took a post-test (the same as the pre-test) and filled in a questionnaire right after the course.

3.2 Subjects

Subjects ($n = 50$) were randomly sampled by the language center from students ($N = 375$) in a technology university in central Taiwan. The 375 students (the defined population) volunteered to sign up for the TOEIC course at the end of the previous semester. For those who failed to enter the course, additional courses were opened in the next semester or the students needed to be on the waiting list since some of the selected applicants might give up their positions. The subjects (42 females; 8 males; age: 20~24) included 46 undergraduates and 4 graduate students. The majors of the students were diverse. They came from 9 different departments. To justify the sample size appropriate for the paired sample *t*-test in the current study, a *priori* power analysis using

statistical computing software G*Power (v. 3.1.7) was performed. Four input parameters (two-tailed; effective size (Note 3) = 0.833; $\alpha = 0.05$; power = 0.9) were entered and the calculated total minimum sample size is 18. The power analysis showed that the current study ($n = 50$) has a statistically appropriate subject size for inferential pair sample t-test. It should be noted that the current study defined the population as the 375 voluntary applicants who signed up for the course. Therefore, results are generalized to neither the whole school nor all the TOEIC course students in Taiwan. The actual size of students who take TOEIC courses in Taiwan, which is unspecified, is beyond the scope of the current study.

In terms of English proficiency, the participants were at CEF-A2 level. The four graduate students, despite their more advanced academic performance, demonstrated similar pre-course language proficiency and are thus placed in the same CEF-A2 category. The course was part of the "102-103 Teaching Excellence and Learning Autonomy" (102-130 TELA), a project sponsored by the Ministry of Education (MOE). It lasted for one month in a summer of the school year in 2013. The total number of teaching hours of the course was 48 hours. The course included two parts, 24 hours for the reading part of the TOEIC test and 24 hours for the listening part of the TOEIC test. The instructor, with a degree of M.A.in translating and interpreting, had over 5 years of experiences of teaching TOEIC in college levels. He also had the Certificate of College Instructors, which was issued by Ministry of Education (MOE). The instructor has ever taken formal TOEICs and is familiar with the test format.

3.3 Textbook/Quiz/Equipment/Syllabus/Course Design

Students in the school can only take the course for one time during their school years. The students took a pre-test before the course, took the one-month intensive course, and took a post-test before the closing of the course. On the day of the post-test, the students filled in a questionnaire for their reflection and suggestion on the course. The pre-test and post-test were the same. The 48-hour TOEIC course was divided into 16 periods (8 periods for listening and 8 periods for reading). Each period was three hours. The class met four times a week (Monday to Thursday). The class met a total of 16 week days. There were no classes on Friday, Saturday and Sunday.

A textbook was used by the instructor. The textbook was published by *Crane Publisher*. The textbook was *New TOEIC*, with CDs, prepared test generators and class slides. The classroom for the course was equipped with audio and video players, and an overhead projector for slides. In the first 20 minutes of each period (excluding the first period), the students took a quiz on the materials taught in the previous period. The quiz materials, prepared by the language center, includes 20 multiple questions that simulate the format and extent of difficulty of formal TOEICs. Overall, the intensive course is well-structured and followed a syllabus designed by the instructor.

3.4 Instruments

Two instruments were used in the study. First, a full-version official TOEIC (labeled as D-02), compiled and examined by a group of professional English teachers ($N = 6$) in the language center, was used as the pre-test and the post-test. In a pilot study ($N = 60$), the D-02 test was administered in another TOEIC course for two times to check the test-retest reliability. To avoid intervening factors, such as improvement in language proficiency, the duration between the first and second test was 4 days, three of which were the weekend when there was no class for the students. The bivariate correlation (Pearson; two-tailed; $\alpha = 0.01$) showed that the test-retest reliability was statistically significant ($r = 0.81$, $p < 0.01$). The result showed a high reliability of the D-02 test items.

Second, a self-designed questionnaire was used to collect students' personal information (gender; department) and qualitative information (satisfaction with the course), including a total of 28 5-Point Likert-types questions (Appendix 1). In a pilot study *(N = 60)* using the Exploratory Factor Analysis (EFA) with a minimum-eigenvalue of 1.0, 7 factors were extracted by principal component analysis. A varimax rotation produced a simple factor structure with seven factors loading heavily on different questions. Each of the seven factors received appreciable loadings (i.e. loadings of more than 0.6) from four questions. The seven extracted factors were labeled respectively, *Articulation* (item 1, 3, 5, 10)*, Textbook* (item 2, 4, 6, 15)*, Administration* (item 7, 9, 11, 13)*, Facility* (item 8, 14, 16, 17)*, Quiz* (12, 18, 19, 27)*, Materials* (item 20, 22, 23, 25)*, and Time* (item 21, 24, 26, 28)*.* To further check the internal consistency of the items under each factor, we calculated the Cronbach's α coefficient, ranging from 0.76 to 0.88. The mean coefficient was 0.81. From the results of both EFA and Cronbach's α, we are sure the items included in the satisfaction questionnaire have high content validity and internal reliability.

3.5 Data Collection

The data collected included two parts. For the first part, the scores (listening, reading and total scores) of the pre-test and post-test were collected. For the second part, the results of the questionnaire were collected.

3.6 Data Analysis

The data collected from the subjects (the scores in the pre-test and post-test/the results of the questionnaire) was analyzed separately. For the scores in the pre-test and post-test, the researcher used paired sample *t*-test (two-tailed; α = 0.05) and descriptive statistics to analyze the data. The aspects of analysis were in accordance with the first research questions. The analysis included the following two aspects:

1) Inferential *t*-test results and descriptive score pattern distributions in pre-test and post-test (total scores, the scores of listening and reading)

2) Descriptive scores/score improvement pattern distributions (total scores, the scores of listening and reading)

For the results of the questionnaire, the researcher conducted an intercorrelational analysis (Pearson product moment correlation) and multiple regressions to present statistical results of students' satisfaction with the TOEIC course to answer the second research question.

4. Results

4.1 Inferential T-Test Results and Descriptive Score Pattern Distributions in Pre-Test and Post-Test (RQ1)

As shown in Table 1, students' improvement in total scores after taking the intensive course was statistically significant, *t*(49) = -5.99, *p* < .001, Cohen's d = 0.3, 95% CI[-70.47, -34.93].

Table 1. Students' TOEIC total scores in the pre-test and the post-test

Score	Pre-test (*n* = 50)		Post-test (*n* = 50)		*t*	*p*	95% CI		Cohen's d	1-β
	M	*SD*	*M*	*SD*			*LL*	*UL*		
Total	467	74	520	101	-5.99	.000	-70.47	-34.93	0.7	0.8

As shown in Figure 1&2, the distribution of students' total scores in pre-test and post-test was slightly different. Before taking the course, the distribution of total scores was skewed to the left. Only several were found to fall on the higher score range of the scale. After the taking the course, the distribution of total scores showed a double-peaking. While some students moved to the higher score range, the others still stayed at lower score range.
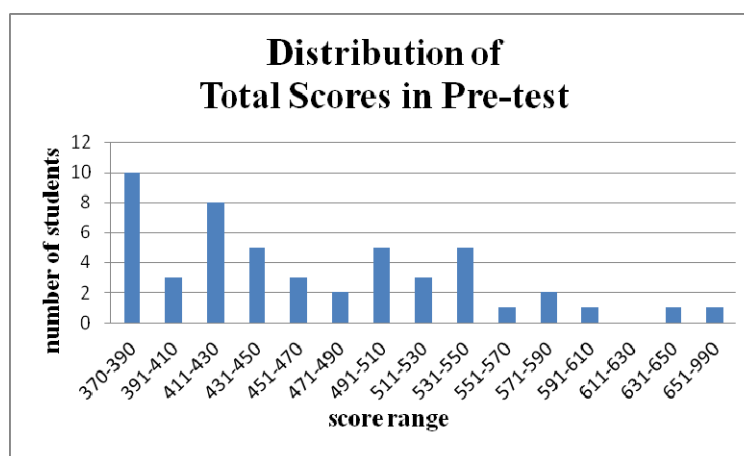

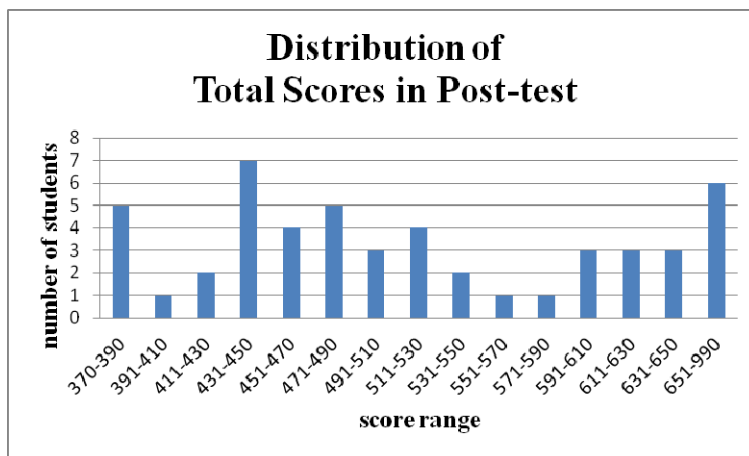Figure 1. Distribution of total scores in pre-test

Figure 2. Distribution of total scores in post-test

As shown in Table 2, students' improvement in listening part after taking the intensive course was statistically significant, $t(49) = -6.4$, $p < .001$, Cohen's d = 0.3, 95% CI[-66.36, -34.63].

Table 2. Students' TOEIC listening scores in the pre-test and the post-test

| Score | Pre-test ($n = 50$) | | Post-test ($n = 50$) | | t | p | 95% CI | | Cohen's d | 1-β |
	M	SD	M	SD			LL	UL		
L	252	52	303	62	-6.4	.000	-66.36	-34.63	.67	.80

Note. L = listening

As shown in Figure 3&4, the distribution of students' listening scores in pre-test and post-test was slightly different. Before taking the course, the distribution of total scores was normally distributed. Only several were found to fall on both the higher and lower score range of the scale. After the taking the course, the distribution of total scores showed a rightward skew. More students moved to the higher-score range. In addition, the number of students on the lower score range dropped sharply.
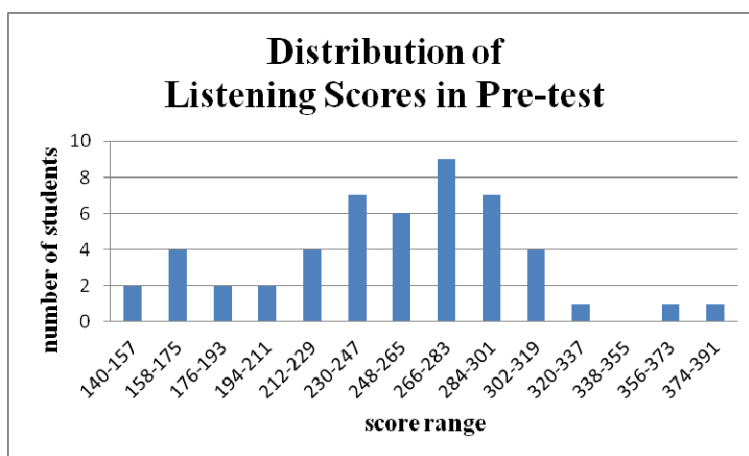

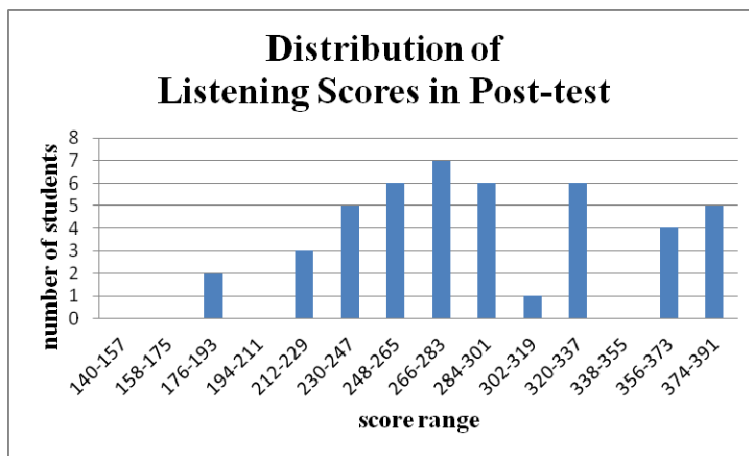Figure 3. Distribution of listening scores in pre-test

Figure 4. Distribution of listening scores in pre-test

As shown in Table 2, students' improvement in reading scores after taking the intensive course was statistically insignificant, $t(49) = -.36$, $p = .719$, Cohen's d = 0.3, 95% CI[-14.43, 10.03]. ($p = 0.719$).

Table 3. Students' TOEIC reading scores in the pre-test and the post-test

Score	Pre-test ($n = 50$)		Post-test ($n = 50$)		t	p	95% CI		Cohen's d	1-β
	M	SD	M	SD			LL	UL		
R	214	47	216	48	-.36	.719	-14.43	10.03	0.34	.80

Note. R = reading

As shown in Figure 5&6, the distribution of students' reading scores in pre-test and post-test was no different. Before taking the course, the distribution of total scores was normally distributed. Only several were found to fall on both the lower and the higher score range of the scale. After the taking the course, the distribution of reading scores barely changed in shape.
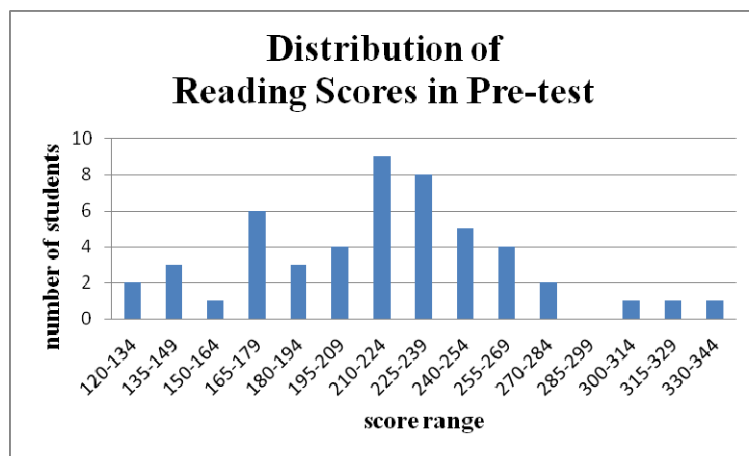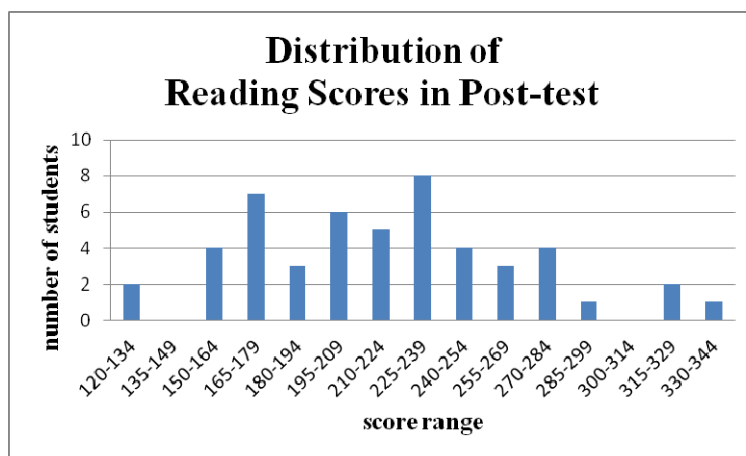

Figure 5. Distribution of reading scores in pre-test

Figure 6. Distribution of reading scores in post-test

4.2 Descriptive Scores/Score Improvement Pattern Distributions (RQ1)

As shown in Table 4, the average improvement of total scores (post-test total score minus pre-test total score) was 52.7. As for outliers, the maximum improvement in total score was 235, while the minimum was -120. The distribution of improvement in total scores was normally distributed.

As shown in Table 5, the average improvement of listening scores (post-test listening score minus pre-test listening score) was 50.5. As for outliers, the maximum improvement in total score was 140, while the minimum was -80. The distribution of improvement in listening scores was a little skewered to the right.

Table 4. Changes in TOEIC total scores

	M	*SD*	Min	Max
Changes	52.7	62.5	-120	235

Note. Min = minimum; Max = maximum
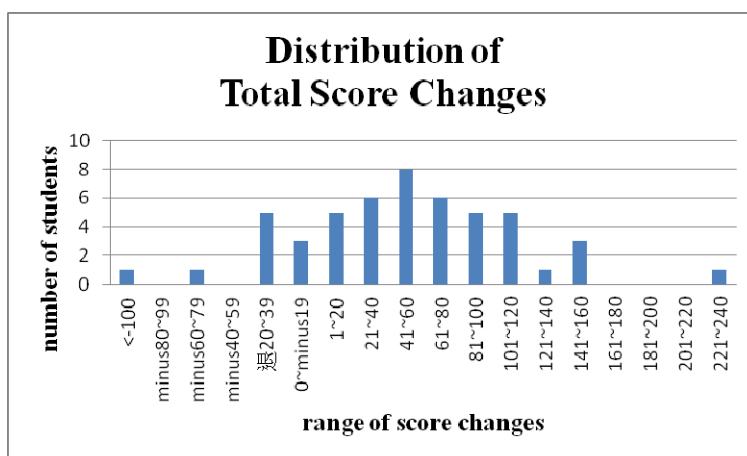
Figure 7. Distribution of total score changes

Table 5. Changes in listening scores

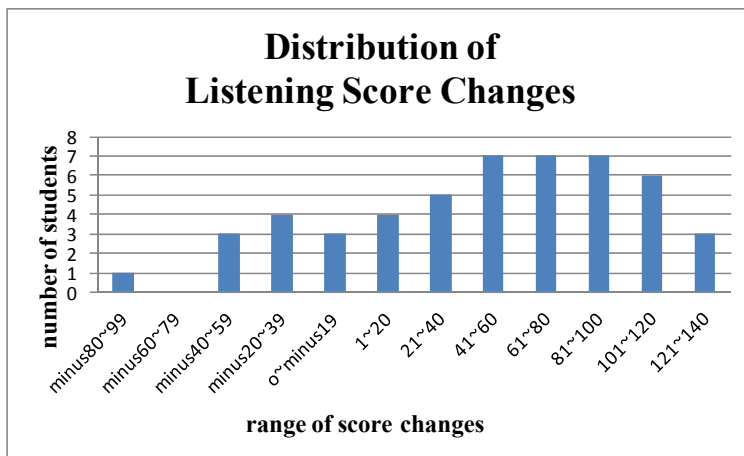|	*M*	*SD*	Min	Max
Changes	50.5	55.8	-80	140

Note. Min = minimum; Max = maximum

Figure 8. Distribution of listening score changes

As shown in Table 6, the average improvement of reading scores (post-test reading score minus pre-test reading score) was 2.2. As for outliers, the maximum improvement in total score was 115, while the minimum was -80. Despite the dramatic outliers, the overall distribution of improvement in reading scores was normally distributed.

Table 6. Changes in reading scores

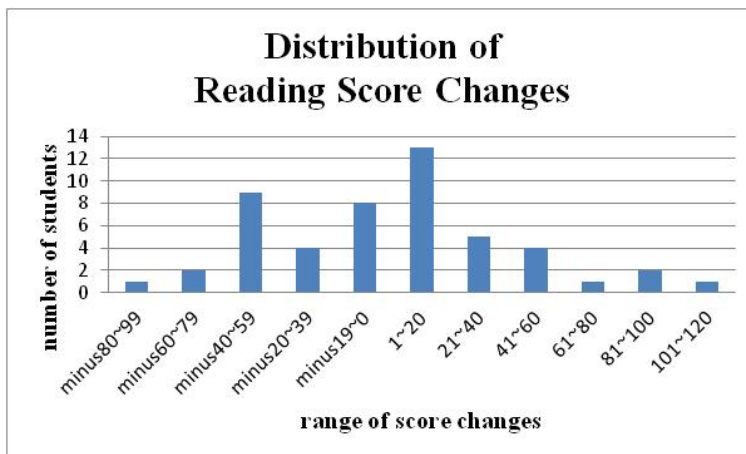|	M	SD	Min	Max
Changes	2.2	43	-80	115

Note: Min = minimum; Max = maximum

Figure 9. Distribution of reading scores changes

Figure 10 gives the comparison of students' improvement in reading and listening scores. The improvement in listening scores was more dramatic than that of reading score. While most of the students' improvement in reading score was moderate, between 0~20, improvement in listening score was relatively significant, with many falling in the score ranges of over 40.
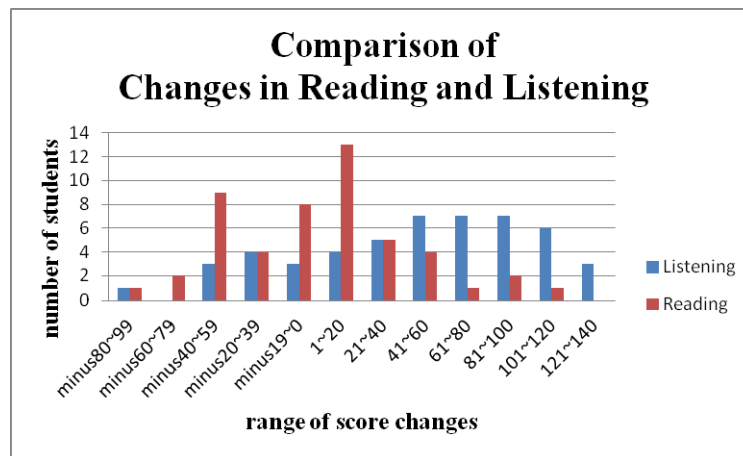
Figure 10. Comparison of changes in reading and listening

4.3 Students' Satisfaction with the Course (RQ2)

As shown in table 7, the overall satisfaction rate was 3.93, which was positively correlated with the satisfaction on articulation ($r = .84$, $p = .00$), textbook ($r = .69$, $p = .00$) and facility ($r = .76$; $p = .00$) and materials ($r = .7$, $p = .00$). The satisfaction on administration, quiz, and time for the course was not statistically correlated with the overall satisfaction.

Table 7. Correlation between satisfaction on articulation, textbook, administration, facility, quiz, materials, time for the course and overall satisfaction (N = 50)

	CWOS	*M*	*SD*
CWOS	1	3.93	.57
Articulation	.84***	4.01	.57
Textbook	.69***	3.98	.55
Administration	.12	1.98	.54
Facility	.76***	2.1	.60
Quiz	.14	1.86	.70
Materials	.70***	4	.53
Time of the course	.08	1.83	.52

Note. CWOS: Pearson correlation with overall satisfaction

***$p = .000$

As shown in Figure 11, the satisfaction on textbook, articulation and materials is relatively high, and is positively and statistically correlated with overall satisfaction. The satisfaction on facility is relatively low but is positively and statistically correlated with overall satisfaction. The satisfaction on time of the course, quiz and administration is relatively low and is not statistically correlated with overall satisfaction.
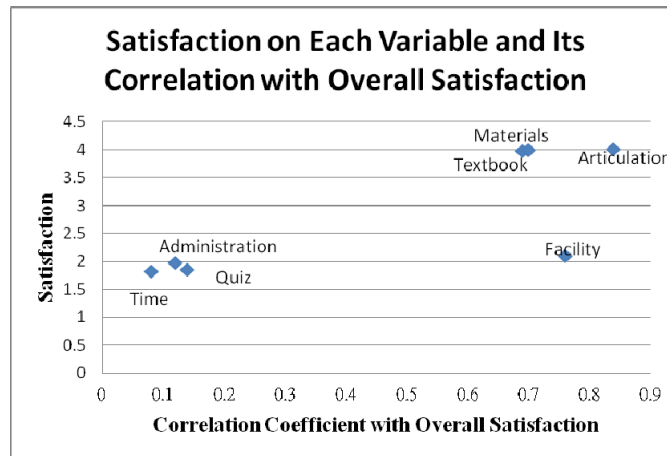
Figure 11. Satisfaction on each variable and its correlation with overall satisfaction

As shown in Table 8, the R^2 is .923, which indicates that the multiple regressions serves as a good model in our analysis of using different variables to predict learners' overall satisfaction on TOEIC course. The adjusted R^2 of .827 indicates that a big portion of variances can be explained by the combined variance of the predictors.

Table 8. Multiple regressions on TOEIC course satisfaction

Predicting Variables							
AR	T	AD	F	Q	TOC	M
B SEB t	*B SEB t*	*B SEB t*	*B SEB t*	*B SEB t*	*B SEB t*	*B SEB t*
OS .578 .561 4.09***	.198 .19 2.38*	ns	.262 .326 4.09***	ns	ns	ns

R^2 = .923

Adjusted R^2 = .827

Durbin-Watson = 1.923

Note. AR: articulation; T: textbook; AD: administration; F: facility; Q: quiz; TOC: time of class; M: materials; OS: overall satisfaction; ns: ns, a non-significant variable removed from final regression models.

p < 0.05; ***p* = .000

The Durbin-Watson of 1.923 further indicates a low extent of collinearity among variables in the model. Among the seven variables, satisfaction on articulation (SE *B* = .561, *t* = 4.09, *p* = .00), textbook (SE *B* = .19, *t* = 2.38, *p* = .00) and facility (SE *B* = .326, *t* = 4.09, *p* = .00) significantly and positively predicts overall satisfaction, while administration, quiz, time of the course, and materials do not serve as statistically significant predictors of overall satisfaction on TOEIC course. Although satisfaction on materials is shown to be statistically and positively correlated with overall satisfaction, it does not work as a statistically significant predictor of overall satisfaction on TOEIC course according to the result of multiple regressions.

5. Discussion

The statistical results presented in the previous section helped answer the two research questions proposed in this study: (1) Can the intensive TOEIC course improve students' scores? What are the score pattern distributions before and after taking the course? (2) What are the factors that are associated with or serve to predict students' overall satisfaction with the course?

For the first research question, we found that students' overall scores improve statistically after the intensive TOEIC course. However, there was an uneven contribution of the course to students' listening and reading scores. The relatively little progress made in the reading scores, as compared with that in listening scores, pinpointed the differential effect of the same course on students' listening and reading scores. This divergence in score gains was similar to Robb & Ercanbrack (1999), who found that students' score gains in reading and listening after taking the TOEIC course were very unstable and highly sensitive to their proficiency levels,

outside-the-classroom practice and satisfaction with the course (Robb & Ercanbrack, 1999). Our finding provided another evidence of this phenomenon. We might, therefore, argue that evaluating the effectiveness of intensive TOEIC courses should be done more carefully. Specifically, evaluating the effectiveness of the course simply by looking at the overall score gains, irrespective of reading and listening, might be misleading. In another perspective, Cunningham (2002) also argued that there was no linear relationship between score gains and real improvement in students' communicative competence. If that is the case, the complex relationship between interpreting score gains and evaluating the effectiveness of TOEIC courses can be even more daunting. However, in the current study, we didn't further examine students' real improvement in communicative competence and its relationship with score gains. This issue is important in providing more diverse perspectives for evaluating the effectiveness of an intensive TOEIC course and is, therefore, recommended for the follow-up study.

With regards to the score pattern distributions, we found that the generally normal distributions in the pre-test became less centralized in the post-test. This is particularly significant in listening and total scores, but not in the reading scores. In both listening and overall scores, more outliers appeared in the higher score ranges and those in the lower-score range disappeared. However, in the reading scores, the distribution remained almost the same. This situation also appeared in the distributions of score improvements. A stark contrast was found between listening and reading score improvements, with the former skewing to the right (higher-score range) while the latter remained in its normal shape. The descriptive score pattern distributions revealed the issue of outliers that inferential statistics couldn't. The inferential statistics could only tell us the general picture of a course but failed to help us identify those who progress or regress significantly as compared to the whole class. Figure 7 and 8 provided evidence that in a single class, the range of students' improvements in overall or listening scores can be as big as 200 points. These outliers, though small in number, might have been ignored in inferential statistics, which focused on the class as a whole. These outliers, nevertheless, are worth our attention since numerous researchers have argued that it is these outliers that help us know the main drawback of the course and the different learning needs of different students. (Harmer, 2010; Larsen-freeman, 2011). Figure 2 and 4 provided evidence that some outliers, either the more advanced or slower learners, might be the ones that a one-fits-all intensive TOEIC course syllabus might ignore. Arguably, given the diversity of the class, some pedagogical concerns are implied pertaining to the use of more individualized and eclectic teaching methods in intensive TOEIC courses. In specific terms, instructors should heed to not only the need of the majority but also the minority, e.g. the outliers who deserve a private tutoring consultancy or a more engaging individualized teaching approach (Carrell, 1985).

For the second research questions, we found some factors that were statistically correlated with students' overall satisfaction on the intensive TOEIC course, such as articulation, textbooks, facility and materials. Two extraneous factors (administration and time of the course) were excluded in the correlational analysis. By breaking up the satisfaction analysis into different categories, we were better informed of the variables that play more important roles in students' satisfaction with the course. The current finding was similar to Lai (2008), who also spotted various factors in affecting learners' response and achievements in a TOEIC course, such as test-taker factors, nature of text, speaker variables, socio-cultural competence and strategic competence (Lai, 2008). In addition to teachers' role, which was specified in both studies, these findings also highlighted the need for course designers to consider extraneous outside-the-classroom factors that might indirectly influence learners' perception of the course, which under the theory of needs-assessment proposed by Savignon (2002) played as a critical role in learners' long-term achievements and motivation (Gardner, 1985). What is different between the current study and Lai (2008) and Savignon (2002) is that in addition to spotting the relationship between various variables, a multiple regression model was proposed indicating the relatively stronger predicators of students' overall satisfaction with the course. The factor of materials (what the instructor teaches in addition to the textbook) was excluded from the regression model. The ultimate three predicting factors extracted under the model were articulation, textbook and facility. This finding is significant in providing statistically reliable information for course designers as well as instructors in interpreting learners' satisfaction. Since these factors serve as positive strong predictors of overall satisfaction, they provide a more concrete direction for improving learners' satisfaction with the course. This major finding substantiated and potentially enlarged the scope of needs-assessment theory by Savignon (2002), which simply dwelled on learners' needs on learning. The proposed model added to numerous studies that placed mounting concerns on the roles of instructors' teaching style, equipment quality and textbook selection in second language learning (Brown, 2010; Harmer, 2010; McGrath, 2002; Tomlinson, 2011). Caution, nevertheless, needs to be exerted in that the result of the satisfaction analysis can hardly be generalized to all the TOEIC course students. Statistically, the sample (n = 50) serves to, at best, reflect the picture of the defined population ($N = 375$) in this paper.

6. Conclusion

6.1 Summary of Findings

The current study found that the intensive TOEIC course is effective in improving students' total scores, but its effect in improving students' scores in listening and reading was differential. The descriptive score pattern distributions further highlighted the issue of outliers (those who progress or regress significantly as compared to the whole class) in TOEIC learning process. In addition, the satisfaction analysis proposed a model that indicated three positive predicting factors for students' satisfaction on the course: articulation, textbook and facility. The findings provided implications in two aspects. First, a more eclectic, individualized teaching approach should be taken to meet not only average but also outlier students' needs in TOEIC courses. Secondly, course designers are provided more concrete direction for improving students' satisfaction on the course.

6.2 Limitation

The current work suffered from the following limitations. First, this is a one-group pre-test post-test study, with no control groups being manipulated. Therefore, generalizing the results to a wider EFL context, e.g. the whole TOEIC learners in Taiwan as a population, should be careful. To further test the effectiveness of intensive TOEIC courses, an experimental study encompasses both the control and the experiment group is called for. Also, the current study only focused on students' satisfaction. Other students' variables, such as pre-class language proficiency, motivation, purposes for taking TOEIC courses were not attended to in the current study. Future studies are needed to examine their relationships with students score gains. Finally, a more comprehensive model needs to be built to establish the relationship between learner satisfaction, pre-class language proficiency, motivation, purposes for taking TOEIC courses and real communicative competence improvement (Cunningham, 2002; Gardner, 1985; Lai, 2008). Such modeling entails potential pedagogical implication for intensive TOEIC course design and implementation.

References

Booth, D. K. (2012). *Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity*. University of Auckland: Applied Linguistics.

Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). NY: Pearson Education.

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL quarterly*, *19*(4), 727-752. http://dx.doi.org/10.2307/3586673

Cunningham, C. R. (2002). *The TOEIC test and communicative competence* (Unpublished doctoral dissertation). University of Birmingham, Birmingham, the UK.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London, GB: Edward Arnold.

Harmer, J. (2007). *How to teach English*. Essex: Pearson Education.

Jee, S., & Kim, H. C. (2013). Understanding English learners preparing for TOEIC and their information technology usage practices in Korea. *International Journal of Multimedia and Ubiquitous Engineering*, *8*(5), 93-104. http://dx.doi.org/10.14257/ijmue.2013.8.5.10

Johnson, D. M. (1992). *Approaches to research in second language learning*. NY: Longman.

Lai, Y. H. (2008). A study on effectiveness of college English-featured courses on TOEIC. *National Kaohsiung Normal University Journal*, *25*, 71-91.

Larsen-Freeman, D. (1986). *Techniques and principles in language teaching*. NY: Oxford University Press.

Lin, C. H. (2012). College student personality traits related to TOEIC and GEPT. *Language Testing in Asia*, *2*, 33-46. http://dx.doi.org/10.1186/2229-0443-2-4-33

McGrath, I. (2002). *Materials evaluation and design of language teaching*. Edinburg: Edinburg University Press.

Miyamoto, T., & Tomoyuki. (2010). Skill-based intensive listening and speaking training for TOEIC®: A preparation course based on underlying theories and instructional principles. *Osaka Women's University Journal*, *4*, 71-83.

Phillips, C. (2006). How can we help students improve their TOEIC scores without "teaching to the test"? *Angelaki: Journal of the Theoretical Humanities, HUMANITIES*, *11*, 55-74.

Phy, S. (2010). *Graduate and employment in the Republic of Korea and Cambodia: Prospects and challenges*. Nordersede: GRIN Verlag.

Richards, J. C., & Rodgers, T. S. (1986). *Approached and methods in language teaching: A description and analysis*. Cambridge: Cambridge University Press.

Robb, T. N., & Ercanbrack, J. (1999). A study of the effect of direct test preparation on the TOEIC scores of Japanese university students. *Teaching English as a Second or Foreign Language*, *3*(4)*, 2-16.

Savignon, S. J. (2002). *Communicative competence: Theory and classroom practice, international second edition*. Singapore: McGraw-Hill Book Co.

Sewell, H. D. (2005). *The TOEIC: reliability and validity within the Korean context.* Retrieved October 10, 2013, from http://72.14.235.104/search?q=cache:jlOewP1wZjgJ:www.cels.bham.ac.uk/resources/essays/Sewell% 2520Testing.pdf+The+TOEIC:+reliability+and+validity+within+the+Korean+context&hl=zh-TW&ct=clnk &cd =1&gl=tw

Tomlinson, B. (2011). *Materials development in language teaching*. Cambridge: Cambridge University Press.

Notes

Note 1. http://www.toeic.com.tw/about_01.jsp

Note 2. http://www.toeic.com.tw/toeic_news_page.jsp?type = 4&gid = 8

Note 3. The effect size is obtained by assuming the hypothesis that the mean difference (score gain) is 50 and standard deviation of difference is 60.

Appendix A

2013 Summer Intensive TOEIC Course After-class Learner Satisfaction Survey

Please read the following statements and circle 5(strongly agree), 4(agree), 3 (neutral), 2(disagree) and 1 (strongly disagree).

1) The teacher's lecture is clear and to the point.	5	4	3	2	1
2) The textbook suits my needs.	5	4	3	2	1
3) The teacher's articulation is logical.	5	4	3	2	1
4) The textbook provides a lot of TOEIC practice questions.	5	4	3	2	1
5) The teacher is good at motivating the class in the lecture.	5	4	3	2	1
6) The textbook is suitable for self-study.	5	4	3	2	1
7) The class assistants are helpful in class preparation and management.	5	4	3	2	1
8) The classroom is well-equipped.	5	4	3	2	1
9) The class assistants are patient and friendly to the students.	5	4	3	2	1
10) The class assistants try their best to suit the needs of the students in all aspects.	5	4	3	2	1
11) The quiz schedule arrangement is great.	5	4	3	2	1
12) The class assistants act as a great go-between the teacher and the students.	5	4	3	2	1
13) The class assistants are patient and friendly to the students.	5	4	3	2	1
14) The acoustic facility of the classroom is great so that I can clearly follow the sentences read aloud during the listening practice.	5	4	3	2	1
15) The textbook is resourceful, clear, and highly readable.	5	4	3	2	1
16) The classroom is equipped with a good-quality microphone. I can follow the lecture much more easily.	5	4	3	2	1
17) The classroom has great air-conditioning.	5	4	3	2	1
18) I think the quizzes are helpful in preparing me for formal	5	4	3	2	1

TOEICs.

19) The difficulty of the quizzes is acceptable for me.	5	4	3	2	1
20) The teacher gives sufficient supplementary teaching materials for us in addition to the textbook.	5	4	3	2	1
21) I am quite adapted to the intensity of course arrangement.	5	4	3	2	1
22) The supplementary teaching materials are useful and not too difficult for me.	5	4	3	2	1
23) The supplementary teaching materials inspire me in several learning aspects.	5	4	3	2	1
24) The intensive course schedule is acceptable for me.	5	4	3	2	1
25) The supplementary teaching materials make up for the drawback of the textbook.	5	4	3	2	1
26) I have a great sense of learning achievement when attending the intensive course.	5	4	3	2	1
27) I am quite adapted to the length of class period.	5	4	3	2	1
28) Overall satisfaction with the course.	5	4	3	2	1

Copyrights