E-Journal of Yaşar University
http://joy.yasar.edu.tr

Journal of Yasar University, 2014 9(36) 6261- 6380

Design and Analysis of Student ECTS Credit Feedback Surveys using Pairwise Comparisons

İkili Karşılaştırma ile Öğrenci AKTS Kredisi Geribildirim Anketlerinin Tasarımı ve Analizi

Raif Serkan Albayrak, Yasar University, Turkey, raif.albayrak@yasar.edu.tr

Abstract: *Student ECTS Credit Feedback Surveys are implemented in order to fine tune ECTS credits of courses which are mainly calculated and allocated by the instructors according to the workload components they declared in the course descriptions. However, accumulated experience shows that the current design of surveys requires immense cognitive load on part of the student. At present, survey questions are posed in such a way that it requires the student to undertake a self-appraisal process regarding the workload of learning activities for a particular course. This article discusses an alternative design for adjusting ECTS credits; a holistic perspective based on pairwise comparisons of workloads of courses. Binary relationships are unfolded using Thurstone scaling and then rescaled to ECTS units through a measurement error model. Monte Carlo simulations are used to demonstrate the robustness of the estimators.*

Keywords. ECTS Credit, Pairwise comparison, Thurstone Scaling, Measurement Error Model

Öz: *Öğrenci AKTS Kredisi Geri Bildirim Anketleri; derslere ait olan ve öğretim elemanları tarafından ders tanımında bildirilen iş yükü unsurları baz alınarak hesaplanan ve belirlenen AKTS kredileri hakkında geri dönüt elde etmek amacıyla uygulanırlar. Ne var ki, uygulamada, anketlerin mevcut tasarımının öğrenci için yoğun bilişsel yük teşkil ettiği görülmüştür. Mevcut anket soruları, öğrencinin her bir ders için gereken öğrenme aktivitelerinin getirdiği iş yüküne dair bir özdeğerlendirme sürecine girmesini gerektirmektedir. Bu makale AKTS kredilerini ayarlamak için alternatif bir tasarım sunmaktadır; derslere bağlı işyüklerinin ikili karşılaştırmasına dayanan bir bütünsel perspektif. İkili ilişkiler Thurstone ölçeği kullanılarak ortaya çıkarılır ve bir ölçüm hata modeli aracılığıyla AKTS birimleri olarak yeniden ölçeklendirilirler. Monte Carlo simülasyonları ile tahminleyicilerin özellikleri raporlanmıştır.*

Anahtar Kelimeler. AKTS Kredisi, İkili Karşılaştırma, Thurstone Ölçeği, Ölçüm Hata Modeli

1. INTRODUCTION

Through the Bologna Accords, Bologna process aims to ensure the comparability in the quality of the standards and the quality of the higher education system by creating a European Higher Education Area with the participation of 47 countries as of Budapest-Vienna declaration (B.V. Deleration, 2010). In order to facilitate this objective Bologna process introduced European Credit Transfer System (ECTS). ECTS is aimed to be a common unit in the design of the curricula that provides a clearly defined system for students and institutions (EC, 2009).

ECTS is now widely recognized standard in design of higher education curricula across the European Union and other collaborating countries. Radical shift from an instructor - oriented system to a student - workload based methodology, has rapidly populated the literature, addressing conceptual and procedural issues with respect to calculations and implementations of ECTS credits.

This article introduces a novel approach for the design and analysis of student's ECTS feedback surveys that are used to regulate whether students are able to perform their tasks in the prescribed period of time. The calculation of ECTS credits rests on three fundamental pillars and student feedback surveys are one of them. The other two pillars; impositional allocation of credits by the instructors and credit allocation by reference to learning outcomes, are only briefly discussed in this article.

In countries where the ECTS is legislated, periodical student surveys are administered in order to measure the actual workload hours it takes the student to complete all planned learning activities of a course. The attention is concentrated on the student's own experience. Survey questions refer to the cells of the tabular form filled by the instructors in the impositional allocation of ECTS credits. Those tabular forms are detailed enough for instructors to report their projections regarding the amount of time and workload required for an average or typical student to accomplish all academic activities such as: Class hours, preliminary studies, homework, presentations, seminars, project and examination preparations, laboratory study and field work. In that way, student surveys enable a substantial scientific tuning process where theoretical workload estimations can be tested against statistics revealed from data. Furthermore, the design of the survey that emulates the ECTS form filled by the instructors provides workload data for each academic activity enabling detailed monitoring of probable mismatches.

However, current practice shows that there are a number of factors that jeopardize the validity of student surveys. According to Karran (2004), historical data collected from student surveys may involve incompatibilities as the course contents typically vary over time and the effect of this factor is multiplied with the changes in staff turnover. In addition, Albayrak and Gurkan (2011), discuss prevailing issues related to the implementations of student surveys. Requiring students, who have already successfully completed the requirements of a course, to fill out these surveys is problematic, since they could be reluctant to provide feedback due to the realization that the result of such action will not alter their academic records. In a similar vein, Albayrak and Gurkan (2011), underscore the effect of cognitive effort of a student spent to recall the amount of time commitment for each and all academic activities of a course throughout the semester and the precision of the data collected from such feedback surveys. In essence administrators of student surveys should expect contamination in data and, thus, need to use appropriate filtering techniques prior to computing statistical estimators. However, the usage of filtering techniques can distort cross-tabulated information regarding workload per academic activity.

Another widely recognized issue related to the implementation and analysis of student surveys, is the fuzziness in the definition of the average student (Tuning Dissemination Conference I, 2008). The time an average student will need to meet the expected learning outcome may vary across student cohorts since workload is affected by many parameters such as; academic background, capacity, abilities, student origin, major and factors related to instruction of the course such as; effectiveness of teaching methods, methods of assessment and language of instruction. If the shape of the distribution of student workloads significantly deviates from the normal distribution, for instance in cases where the distribution is multimodal or skewed, then the conceptual average student is not adequately or fully represented by its statistical estimator.

The quality of the survey data, and in return, the quality of the statistical estimators, depends to a large extent, on the design of the student ECTS feedback survey. Experience shows that the design of the survey should enable students to report their workloads undergoing minimum cognitive effort. This article develops a survey design that uses ordinal measurement through binary comparisons of workloads of courses for estimating ECTS credits. In each question, the total workload of two courses that the student is enrolled in, is being compared. It is argued that students are more successful in ordinal workload comparisons rather than reporting cardinal workload values of each course. Binary comparisons require a holistic perspective, whereas reporting cardinal values requires recalling and potentially computing working hours for each and every academic activity.

Survey designs based on binary comparisons is efficient compared to current survey design in terms of number of questions as well. Current surveys ask 10 to 15 questions per course, amounting to 50 to 60 questions for a student enrolled in six courses. On the other hand, a binary comparison - based survey requires answering at most 20 questions since the answer of some questions unfolds through the transitivity property of the order relationship.

In binary comparison - based designed student ECTS feedback surveys, every student compares the workload commitment of two courses. These surveys include all courses that the student is enrolled in per semester. When binary comparisons of courses are consolidated across students, workloads can be projected onto a scale using various techniques including Thurstone's Scaling Algorithm (Thurstone, 1927) and the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952; Luce, 1960). The basic premise of developing a scale from binary comparison data can be explained using a simple example: Suppose that 90% of the students say course A required more workload (hours) than course B and the remaining 10% say course B required more workload than course A[1]. Now suppose, further, that 55% say course B required more workload than course C. Then we should conclude that the separation between required workloads of two courses, A and B, on a scale of ECTS credits, is greater than the separation between B and C on the same scale. This logic may as well translate to talk about 'distances on a scale of ECTS credits.

It is known that Thurstone scales are arbitrary; meaning that they are essentially unitless. It is the aim of this article to revert Thurstone scales, obtained from binary comparisons, to actual ECTS units. The next section briefly discusses Thurstone scaling algorithm related to binary comparison- based designs of ECTS feedback surveys. A simulation is used to demonstrate the extraction procedure of Thurstone scales. Various probable

[1] For Thurstone's initial formulation, while binariness is crucial it is as well possible to extend the framework to a multinomial case.

singularities are discussed. These scales are then transformed into ECTS units using a measurement error model. Coherence of instructor's imposed ECTS credits is tested against survey data.

2. THURSTONE SCALING

In the simplest setup, we assume a group of students enrolled in an identical list of courses. This strong assumption can effectively be relaxed at the cost of introducing additional complexity. We further assume that the student workload of a course is normally distributed. Figure 1 depicts the workload distributions $f(X_A)$ and $f(X_B)$ of students that are enrolled in courses A and B.

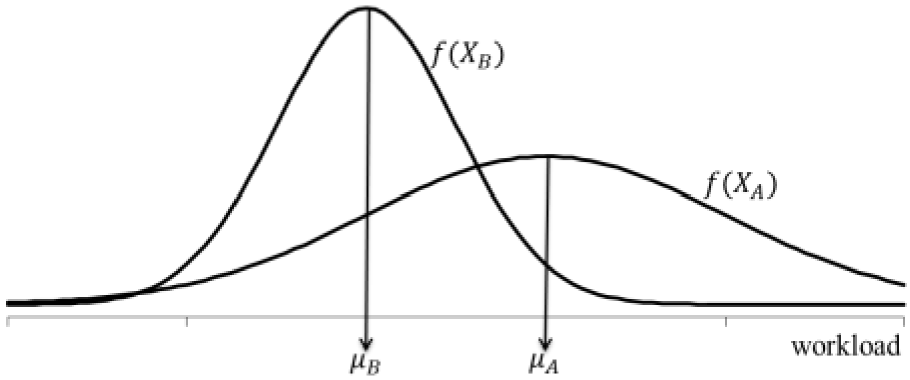


Figure 1. Student workload distributions of two courses

In Figure 1, μ_A and μ_B are unknown average ECTS credits of courses A and B, respectively. Actually every single judgment in the form of; 'A is greater than B' or 'B is greater than A' in terms of workload, involves two variates from two distributions. As shown in Figure 1, when a student spends more hours on course A compared to course B, $X_A^s - X_B^s$ is positive for that student and negative otherwise, where the superscript stands for the particular student.

For the differences in workloads of two courses we should have distribution as shown in Figure 2. In this figure the base line, this time, represents experienced workload differences in ECTS units ($X_{A-B} = X_A - X_B$) and corresponing ordinate represents relative frequencies of such actual experiences. To the right of the origin these differences are positive. The most common experience, which is assumed to represent the average students' workload difference, happens to be the highest. Left of the origin represents the cases for which (X_{A-B}) are negative. Dispersion around the mean of this curve is a function of individual dispersions of the curves in Figure 1.

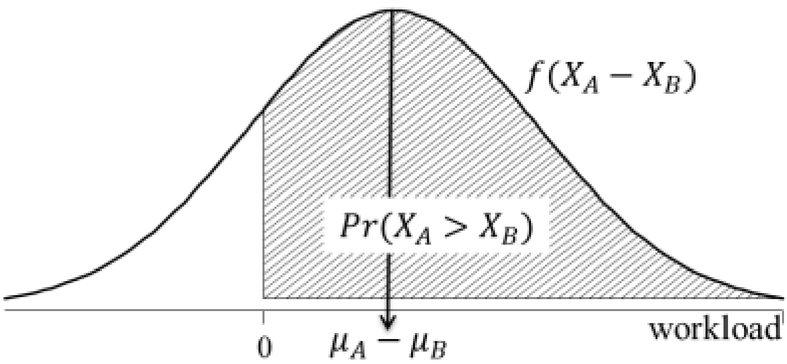


Figure 2. Workload discrepancy distribution of two courses

From this point forth, setup relates observation and theory. Observed $\pi_{A>B}$ is the proportion of students who claimed that the workload of course A is more than the workload of course B. This proportion also corresponds to the shaded area in Figure 2. Accordingly, $\pi_{B>A}$ is the area of the unshaded region. Hence;

$$\pi_{A>B} = \Pr(X_A > X_B) = \Pr(X_A - X_B > 0) = F\left(\frac{\mu_{A-B}}{\sigma_{A-B}}\right) = F(Z_{A-B}) \quad \text{Eq. (1)}$$

$$F^{-1}(\pi_{A>B}) = Z_{A-B} = \frac{\mu_{A-B}}{\sigma_{A-B}} \quad \text{Eq. (2)}$$

Here Z_{A-B} is the value that corresponds to probability $\pi_{A>B}$ in the standard cumulative normal distribution. Also,

$$\mu_{A-B} = \mu_{\text{A}} - \mu_B \quad \text{Eq. (3)}$$

$$\sigma_{A-B} = \sqrt{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}} \quad \text{Eq. (4)}$$

This idea constructs an equation system: $Z_{i-j} = m_i - m_j$ for all binary comparisons between all courses. A solution set $\{m_i | i = 1, \dots, L\}$ defines a Thurstone scale. Next section discusses solution procedures through a simulation example. Various properties of Thurstone scales are discussed at length in Torgerson (1958) and Kenneth (2003).

3. FROM THURSTONE SCALES TO ECTS CREDITS

The workload estimators obtained through the previously explained Thurstone scaling procedure would satisfy probabilistic requirements, $(\hat{m}_i - \hat{m}_j) \cong \pi_{i>j}$. However, since the estimators are standardized, the estimated values would be substantially different from the actual ECTS values. Yet, ECTS credits are themselves arbitrary; they are the quotients of actual workload hours and institutionally decided multipliers (a number between 25 and 30). Since each binary comparison is carried over to workloads, Thurstone scales and ECTS credits have to be two measures of the same concept. As an analogy, Celcius and Fahrenheit are two scales measuring the same concept: temperature. One can relate these scales using a simple linear transformation. In fact, the use of any type of interval scales measuring the same concept, necessitates linear transformation. Thus, all we need to do is to find the *scale equalizer linear transformation* between the instructor's prediction on the ECTS credit of the course and the Thurstone scale calculated from the survey data. Now, if Y_i is the instructor's ECTS credit prediction and m_i is the Thurstone scale corresponding to the same course, the parameters of this linear transformation can be found by solving the equation;

$$Y_i = \alpha + \beta m_i \quad i \in \{1,2, \dots, N\} \quad \text{Eq. (5)}$$

Equation above may appear like an ordinary regression model, with the predictor variable m_i, but it is not for at least two reasons: First, the regressor variable $\hat{m}_i$ () is itself a prediction of a previous model (Thurstone setup) and therefore subject to random error. Second, the dependent variable is coming from another measurement effort; it is the workload prediction of the instructor regarding the average workload and is clearly subject to error. It is known that when there are errors in variables, estimations based on the standard assumptions of ordinary regression leads to inconsistent estimates (Fuller, 1987). A proper approach to compute the parameters of the equation is to use the measurement error model methodology. See Fuller (1987) and Cheng and Van Ness (1999) and references cited herein for the theory and some applications of measurement error models. In order to obtain appropriate estimators, measurement error models require external information about the error variances and the correlation between the errors. For instance, if the reliability ratio of survey data is known or given, consistent estimators of model parameters can be computed. In the next section, a solution where the reliability ratio is given is presented. For other solution variants, refer to Fuller (1987).

In order to test the hypothesis that instructors' and student survey workload estimations are compatible with each other, it is possible to use confidence intervals around the model fit. If an instructor's workload estimation falls outside a pre-determined confidence interval, this serves as evidence for an inconsistency between the workload perceptions of two immediate stakeholders of education.

4. SIMULATION

Using a simulation run, we next show how to use and extend this setup as a decision making tool for comparing instructors' imposed ECTS credits with student ECTS credit feedback survey results. We also discuss solutions for probable singularities that may appear in practice.

Simulation run assumes 50 students from the same study and degree programme enrolled in five courses. These courses can be thought to constitute one semester study program. In practice, there might be several situations that fall outside this assumption such as; students of the same degree programme might be enrolled in different electives and there might be students outside the degree programme taking these particular courses. While the solution procedure suggested in this article can be extended to handle these types of situations, the treatment and indexing can get quite complicated. Therefore we restrict the simulation to the most basic scenario.

Workloads of courses are independently normally distributed with parameters displayed in Table 1. The unit of the parameters is ECTS credits. Standard deviations are chosen in such a way that there is no apparent dependency within means[2].

Table 1. Moments of workload distributions of courses

Course	Mean μ_i	Standard Deviation σ_i
C_1	3	2
C_2	6	3
C_3	4	3
C_4	7	5
C_5	10	4

A binary comparison decision of a student between any two courses, say course 1 and course 2, is simulated by drawing two random variates, C_1 and C_2, from the workload distributions of each course and then compared. Although the average workload value of course 2 is considerably bigger than the average workload value of course 1, workload ordering, for more than 20% of the students, is expected to be: $C_1 > C_2$. Each student performs all binary comparisons by comparing five randomly drawn variates, one for each course. The lower triangle in Table 2 shows ordering counts obtained with this method. In this run, 46 out of 50 students reported $C_2 > C_1$ in terms of workload. Therefore,

$\pi_{C_1>C_2} = 8\%$ and $F^{-1}(\pi_{C_1>C_2}) = Z_{C_1>C_2} = -1.405$, is in the upper triangle.

Table 2. Observed $Z_{C_i>C_j}$ (above diagonal) and sum of $C_2 > C_1$ values (below diagonal). The rightmost column depicts predicted Thurstone scales.

	1	2	3	4	5	Thurstone Scales
1	-	-1.405	-0.202	-0.706	-2.054	-0.873
2	46	-	0.524	0.000	-0.915	0.203
3	29	15	-	-0.412	-1.555	-0.458
4	38	25	33	-	-0.643	0.095
5	49	41	47	37	-	1.033

4.1. Removing Singularities

In the simulation run, for all of the students $C_5 > C_1$ in workload, and this situation creates singularities since both $\mathrm{F}^{-1}(0) = -\infty$ and $\mathrm{F}^{-1}(1) = \infty$. Gulliksen (1956) omits all such 0/1 entries and computes Thurstone scales with the remaining data. However the application of such would cause losing strong preference information of two courses with respect to each other. Another widely adopted solution is substituting 0/1 with $\frac{1}{S}$ ve $1 - \frac{1}{S}$, where S is the number of students that responded to the survey. This value is 50 in this particular simulation. In Table 2, this adjustment is made in the bottom left cell.

[2] Simulation is coded in R (R Core Team, 2013) and for replicability we set the seed to 59. The reason to select this seed is to demonstrate some singularities that may appear in practice.

4.2. Weighted Regression

Thurstone scales $\{m_{C_1}, m_{C_2}, m_{C_3}, m_{C_4}, m_{C_5}\}$ are displayed in the rightmost column of Table 2. Although these values can simply be calculated by taking row averages, in cases where the observed proportions are considerably different from 50%, the model needs to be re-set as a weighted least square regression model so as to equalize error variances (Neter, Wasserman and Kutner, 1996). In such case, we may define the regression equation using the following coding:

$$F^{-1}\left(\pi_{C_i>C_j}\right) = m_{C_i} - m_{C_j} \quad \text{for } i \neq 1 \qquad \text{Eq. (6)}$$

$$F^{-1}\left(\pi_{C_1>C_j}\right) = -2m_{C_j} - \sum_{k\neq\{1,j\}} m_{C_k} \text{ for } i = 1 \qquad \text{Eq. (7)}$$

Here, the second equation forces the sum of the estimators to zero. As a result, any statistical software package that can restrict the equation to pass through the origin, may be used to solve the system. Thus, we have chosen to use R Software (2013) with the weights $w_{ij} = \sqrt{s_{ij}(\pi_{C_i>C_j})(\pi_{C_j>C_i})}$, where s_{ij} is the number of students who compared courses C_i and C_j. Table 3 reports un-weighted and weighted regression estimators and their standard errors.

Table 3. Predicted Thurstone Scales and predicted Weighted Thurstone Scales with standard errors

m_{C_i}	Th. Scales	Stand. Errors	Weighted Th. Scales	Stand. Errors
m_{C_1}	-0.873	-	-0.800	-
m_{C_2}	0.203	0.102	0.169	0.096
m_{C_3}	-0.458	0.102	-0.447	0.094
m_{C_4}	0.095	0.102	0.096	0.089
m_{C_5}	1.033	0.102	0.983	0.112

4.3. Measurement Error Model

In order to relate instructors' estimation of workloads and Thurstone scales calculated from survey data, we set up a measurement error model.

$$Y_{C_i} = \alpha + \beta M_{C_i} + \varepsilon_{C_i} \qquad \text{Eq. (8)}$$

$$m_{C_i} = M_{C_i} + \delta_{C_i} \qquad \text{Eq. (9)}$$

Measurement error model assumes that both Thurstone scales calculated from survey data and survey data itself are erroneous measurements. In equation (8), Y_{C_i} is the instructors' report of average workload in ECTS units. However, both due to survey data and Thurstone scale calculations, students' standardized perception of workload M_{C_i} can only be observed as m_{C_i} because of error, δ_{C_i}. Disregarding systematic errors in the observables can lead to biased and inconsistent parameter estimates, which may confound theoretical conclusions (Fuller, 1987). Distributional assumptions about the parameters of the error measurement model are listed below.

- $\varepsilon_{C_i} \sim N(0, \sigma_\varepsilon^2)$, $\delta_{C_i} \sim N(0, \sigma_\delta^2)$. Also ε and δ are uncorrelated and mutually independent. Thus, $\left(\varepsilon_{C_i}, \varepsilon_{C_j}\right) = 0$, $i \neq j$ and $Cov\left(\delta_{C_i}, \varepsilon_{C_j}\right) = 0$ for all *i* and *j*.

- Variances of *M* and *m* are σ_M^2 and $\sigma_m^2 = \sigma_M^2 + \sigma_\delta^2$ respectively.

These assumptions relate population and sample moments shown in Table 4.

Table 4. Population moments and their sample equivalents

Population Moments	Sample Moments	Sample Values
μ_m	$\bar{m}$	~0
μ_Y	$\bar{y}$	6.000
σ_m^2 $(= \sigma_M^2 + \sigma_\delta^2)$	s_m^2	0.461
σ_Y^2 $(= \beta^2\sigma_M^2 + \sigma_\varepsilon^2)$	s_y^2	7.500
σ_{Ym} $(= \beta^2\sigma_M^2)$	s_{Ym}	1.831

For identifiability, measurement error models require external information or assumption for some of the model parameters (Fuller, 1987). Such information includes the knowledge of at least one of:

1. α in equation (8)
2. error variance σ_ε^2 in equation (8)
3. error variance σ_δ^2 in equation (2)
4. reliability ratio $\kappa = \frac{\sigma_M^2}{\sigma_m^2}$
5. ratio of error variances $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$

For the purpose of this article, controlling the reliability ratio encompasses combined accuracy of survey and Thurstone scaling measurement of student workload. Since it is known that Thurstone scaling has a high reliability ratio (Brown and Peterson, 2009), most of the conjectured inaccuracy in measuring student workload can be projected onto survey data. Yet, through repeated applications of surveys, reliability ratio can be calculated in a straightforward manner. Furthermore, as shown in proceeding pages, prudent conjectures on reliability ratio can relax confidence intervals around model predictions, which in return relaxes incompatibilities between instructor's and students' workload reports for a course.

Given reliability ratio κ, the model parameters can be approximated by substituting sample moments (Table 5). In the simulation we let $\kappa = 0.970$ be the coefficient of determination (R^2) of Thurstone scales.

Table 5. Model parameters and predicting sample moments

Model Parameters	Parameters i.t.o. Sample Moments	Simulation Values
β	$\frac{S_{Ym}}{\kappa S_m^2}$	4.097
α	$\bar{Y} - \beta\bar{m}$	6.000
σ_M^2	$\frac{s_{Ym}}{\beta}$	0.358
σ_δ^2	$s_m^2 - \sigma_M^2$	0.011
σ_ε^2	$s_Y^2 - \beta^2\sigma_M^2$	~ 0

Therefore the parameters are computed as in the third column of Table 5. The equalizer linear transformation then takes the form,

$$Y_{C_i} = 6.000 + 4.097 m_{C_i}$$

Table 6 lists instructors' imposed ECTS credits and forecasted ECTS credits. In this simulation it is assumed that instructors' impositions are true in reality. Forecasted values are obtained from a binary comparison based design of student ECTS feedback survey, where a measurement error model follows Thurstone scaling procedure.

Table 6. Instructor's imposed ECTS credits and outputs of the binary comparison scaling model followed by error measurement model

Course	Instructor's ECTS Projection	Model ECTS Prediction
C_1	3	1.46
C_2	6	6.54
C_3	4	3.46
C_4	7	6.96
C_5	10	11.57

4.4. Decision Making

Student surveys are used as a mean to tune instructor's projections on average student workload. Therefore we define the decision making criteria as a hypothesis test where we assume instructor's projection match with student survey data. One further assumption in the decision making process is that *more than half* of the instructors' projections are actually true. Otherwise the error measurement model would incorrectly scale the workload values.

In order to conduct the hypothesis test, we need to evaluate confidence intervals around the regression line elicited from the measurement error model. As it is customary for such models, we have used the method of stochastic differentials in the derivation of confidence intervals. Thus, the variance around the regression line can be computed as:

$$Var(\hat{y}|m=m_0) = Var\big(\bar{y}-\beta(m_0-\bar{m})\big) = \left\{\frac{\partial\hat{y}}{\partial\bar{y}}\right\}^2 Var(\bar{y}) + \left\{\frac{\partial\hat{y}}{\partial\beta}\right\}^2 Var(\beta) + \left\{\frac{\partial\hat{y}}{\partial\bar{m}}\right\}^2 Var(\bar{m})$$

$$+2\left\{\frac{\partial\hat{y}}{\partial\bar{y}}\right\}\left\{\frac{\partial\hat{y}}{\partial\beta}\right\}Cov(\bar{y},\beta) + 2\left\{\frac{\partial\hat{y}}{\partial\bar{m}}\right\}\left\{\frac{\partial\hat{y}}{\partial\beta}\right\}Cov(\bar{m},\beta) + 2\left\{\frac{\partial\hat{y}}{\partial\bar{y}}\right\}\left\{\frac{\partial\hat{y}}{\partial\bar{m}}\right\}Cov(\bar{y},\bar{m})$$

$$= Var(\bar{y}) + \beta^2 Var(\bar{m}) + (m_0-m)^2 Var(\beta) - 2\beta Cov(\bar{x},\bar{y}) \qquad \text{Eq. (10)}$$

Where; $Cov(\bar{m},\beta) = Cov(\bar{y},\beta) = 0$ and the equations of the variances and covariances can be found as in Gillard and Iles (2006, p.10-11), with the exception that we have used degrees of freedom (L-1) as divisor in the computations for obvious reasons. Using simulation values,

$$Var(\hat{y}) = 0.047 + 0.518(m_0-\bar{m})^2$$

Then the 95% confidence interval around the error regression model can be represented by using the square root of the above expression multiplied by $t_{0.975,(L-1)} = 2.776$. The resulting graph is displayed in Figure 3. The confidence intervals bound the rejection regions for the null hypothesis in decision making. If any point lies outside of this region, one should conclude that there is statistically significant evidence against the null hypothesis for that particular point where the null hypothesis is a congruence between instructor's imposition and survey data.

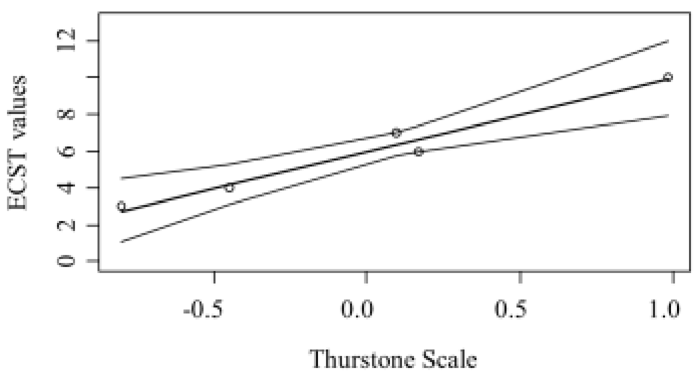


Figure 3. Confidence Intervals around the regression equation

Although it would not impair the calculation of confidence intervals, values too low for reliability coefficient κ would cause σ_{ε}^{2} to be negative. In practical applications this condition can be set as an admissibility constraint; a lower bound for the joint reliability of survey data and Thurstone scale computations. More specifically, reliability constraint should be greater or equal to the square of correlation between independent and dependent variables; $\kappa \geq \rho_{Ym}^{2}$. Nevertheless, when a reliability coefficient below this lower limit is used, it is still possible to compute relaxed confidence intervals (Figure 4).

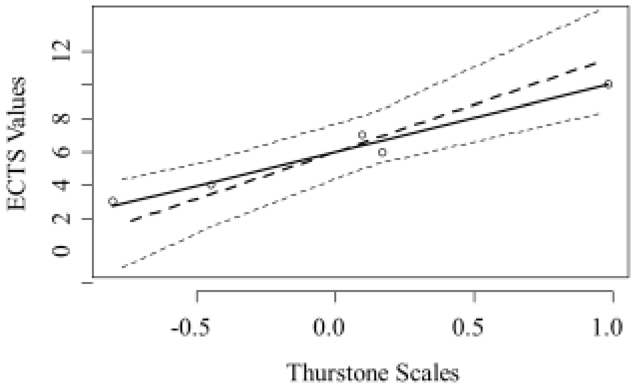
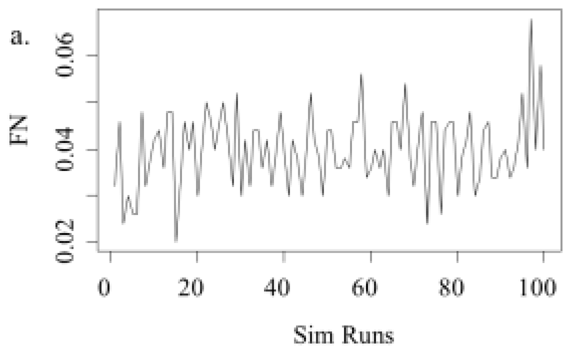


Figure 4. When reliability coefficient is decreased to 0.70, it is still possible to compute a regression line and corresponding confidence intervals. Dashed and solid regression lines correspond to solutions for $\kappa = 0.70$ and 0.97 , respectively.

4.5. Robustness

Any decision making model under uncertainty is subject to false negatives (FN) and false positives (FP). In this context, the false negative rate is the probability of rejecting an instructor's imposed ECTS credits for a course where it is actually the correct value. The false positive rate is the probability of failure to reject instructor's imposition when it is actually wrong. For a robust decision making model, these probabilities should be as low as possible. In a hypothesis testing procedure the probability of FN is kept under control. That is why 95% confidence intervals around the regression life of error measurement model are used. On the other hand, FP probability, which is also equal to one minus the power coefficient of a test, depends on the true workload distribution of students and can therefore only be simulated.

In order to estimate FN and FP rates, 100 Monte Carlo simulation runs for each rates, with different initial parameters per run, are used. These simulations are further replicated 100 times. Figure 5a displays FN rates. FN rates are within the expected 5% range. In order to calculate FP probabilities, instructor's projection for one randomly selected course deviated from students' ECTS average value. FP rates, or the rate of failure to capture the deviation between instructor's and students' ECTS values, are around 16%, which implies a power ratio of 84% for the decision making process (Figure 5b).



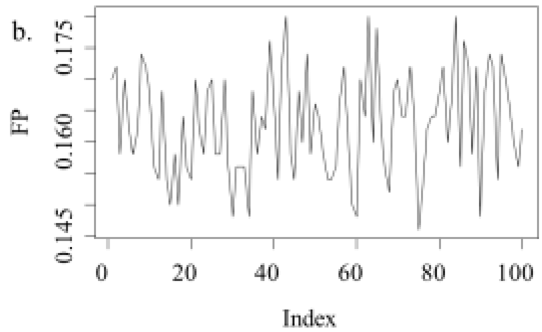


Figure 5. (a) False Negative Rates and (b) False Positive Rates in the Monte Carlo Simulation

4.6. Increasing the Reliability Ratio

Any survey data, in particular mandatory survey data, is subject to contamination. However the discrimination between contamination and variation should be properly made. Rasch literature introduces a misfit indicator called infit value, that can handle this task from a statistical perspective (Curtis, 2004). The idea is to compare a particular student's binary comparisons with the consensus and then compute residuals:

$$R_{j,A,B} = S_{j,A,B} - \hat{\pi}_{A,B} \mid S = 1,0 \qquad \text{Eq. (11)}$$

Where, $S_{j,A,B}$ is Student j's workload binary comparison decision for courses A and B. If for that student the workload of course A is greater than the workload of course B, then $S_{j,A,B} = 1$, otherwise $S_{j,A,B} = 0$. Furthermore, $\hat{\pi}_{A,B} = \mathrm{F}(x_A - x_B)$, and $F(.)$ is the cumulative standard normal distribution. Check that if student workload ordering is in pace with the consensus then residual is less than 0.5. When this residual is standardized, we arrive at a normally distributed statistic for a single binary comparison.

$$z_{j,A,B} = \frac{R_{j,A,B}}{\sqrt{\pi_{A,B}(1 - \pi_{A,B})}} \qquad \text{Eq. (12)}$$

Whenever the squares of these statistics is calculated for all binary comparison for that particular student and summed up, we obtain $z_j^2 = \sum_{\substack{A,B \\ A \neq B}} z_{j,A,B}^2$ value that has Chi Square distribution with [L(L-1)]-1 degrees of freedom. Using this method, it is possible to compute a p-value that can be used to reject a student's complete set of binary comparisons.

Another possible course to follow in order to increase reliability of survey data is to make use of inherent transitivity in binary workload comparison relationship. If, for a particular student, workload of course A is more than workload of course B, and furthermore if workload of course B is more than workload of course C, then for that student, workload of course A is necessarily more than workload of course C. Thus it is possible to check binary comparisons of students for transitivity property and eliminate or restrict student's entries.

5. CONCLUSION

ECTS credit student feedback surveys are essential elements for fine - tuning ECTS credits and a proper implementation of Bologna Process. There are multiple factors that jeopardize the implementation, validity and reliability of these surveys. This article introduces a novel survey design based on the workload comparisons of courses. Rather than asking the students to report direct measures (cardinal values) of the workloads, students compare the workloads of courses in a binary fashion. This new design requires a new model for analysis of the survey data obtained in this way. The model is based on two tiers: Thurstone scaling and measurement error model. First tier arbitrarily scales the workloads of courses using binary comparison matrix whereas second tier reverts the scale to ECTS credits. Robustness of the model is tested using simulation runs.

This type of survey implementation not only remedies weaknesses of usual ECTS feedback surveys, but also overcomes many complications including; the difficulty in the interpretation of the average student workload, high cognitive burden required to recall cardinal working hours committed for an academic activity, and data contamination. This new model also endows the administrators of the survey the control of the reliability ratio of the overall process through which prudent claims about incompatibility between instructor projections and student survey results can be made.

The simulation example used in the paper demonstrates an ideal situation, where all students of a degree programme and, hence, a considerably homogenous population, are enrolled on the same list of courses. However in reality, students of the same degree programme are usually enrolled in different courses as electives and there are also students from other degree programmes taking many such courses. In this case the binary comparison matrix may not behave as nice as it has been idealized in the simulation example. Still, binary comparisons can be unfolded using the techniques discussed in this paper with minor modifications.

Our conjectures on the proposed ECTS survey design and its analysis requires empirical justification. We conjectured that our design requires less cognitive effort to complete the survey. Another assumption was that data collected through binary comparisons are less contaminated when compared to current design. In order to test these conjectures, a lengthy field work is necessary which is a work in progress.

REFERENCES

Albayrak, R.S., & Gurkan, B. (2011). Calculation of ECTS Credits: Case of Yaşar University (In Turkish*)*. *Hacettepe University Journal of Education*, 41, pp.13-23.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3/4), pp.324-345.

Brown, T. C., & Peterson, G. L. (2009). An enquiry into the method of paired comparison: reliability, scaling, and Thurstone's Law of Comparative Judgment. US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Cheng, C. L., & Van Ness, J. W. (1999). *Statistical regression with measurement error* (Vol. 6). London: Arnold.

Curtis, D. D. (2004). Person Misfit in Attitude Surveys: Influences, Impacts and Implications. *International Education Journal*, *5*(2), pp.125-143.

Declaration, B. V. (2010). Budapest-Vienna Declaration on the European Higher Education Area. (reached from web site on 04.12.2012) http://www.ond.vlaanderen.be/hogeronderwijs/Bologna/2010_conference/documents/Budapest-Vienna_Declaration.pdf

EC–European Commission. (2009). ECTS Users' Guide. *Office for Official Publications of the European Communities*, 60.

Fuller, W. A. (1987). Measurement error models. *Wiley Series in Probability and Mathematical Statistics.* New York: Wiley

Gillard, J. W., & Iles, T. C. (2006). Variance covariance matrices for linear regression with errors in both variables. *Cardiff University School of Mathematics Technical Report*.

Karran, T. (2004). Achieving Bologna convergence: is ECTS failing to make the grade?. *Higher Education in Europe*, *29*(3), pp.411-421.

Kenneth, E.T. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Luce, R. D. (1960). Individual choice behavior, a theoretical analysis. *Bull. Amer. Math. Soc.*, 66, pp.166-168.

Neter, J., Wasserman, W., & Kutner, M. H. (1996). *Applied linear statistical models*. Vol. 4. Chicago: Irwin.

R Core Team. (2013). R: A Language and Environment for Statistical Computing.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4), 273.

Torgerson, W.S. (1928). *Theory and Models of Scaling*. Wiley, Oxford.

Tuning Dissemination Conference I, 2008, (reached from web site on 04.12.2012) http://www.unideusto.org/tuningeu/images/stories/presentations/tuning%20disemination%20conference%20I/Conclusions.pdf