PEER EVALUATION OF TEAMWORK AND INDIVIDUAL STUDENT ACHIEVEMENT

ABSTRACT
The purpose of this study is to determine whether peer evaluations can effectively contribute to student course grades; in other words, whether student individual performance indicators correlate with the assessment by peers. 390 undergraduate students from four different instructional design courses were the subjects of the study. Students worked in groups of 2 – 7 to complete team projects. Students also did peer evaluations of their team members regarding the project they completed, rating the performance in the project of their peers and themselves. Peer evaluations were on a scale of 0 to 100 and were averaged for each individual student to produce a mean peer evaluation score. This score included a self-score and scores given by the peers. ANVOCA results showed that peer evaluations can predict the student midterm scores; however this was true only on a limited capacity. Students' exam scores also differed across the courses.
Keywords: Teamwork, Group Study, Student Achievement, Engagement, Peer Evaluation, Assessment

GRUP ÇALIŞMASINDA AKRAN DEĞERLENDİRME VE BİREYSEL ÖĞRENCİ BAŞARISI

ÖZET
Bu çalışmanın amacı, akranlar tarafından yapılan değerlendirmelerin öğrenci notuna etkili bir şekilde katkı sağlayıp sağlamadığını tespit etmektir. Diğer bir deyişle, çalışma, akran değerlendirmelerinin öğrencinin performans göstergeleri ile olan korelasyonunu incelemektedir. Çalışmanın grubunu öğretim tasarımı ile ilişkili dört farklı derse yazılmış olan 390 öğrenci oluşturmuştur. Öğrenciler, 2 ile 7 arası kişiden oluşan gruplarda çalışarak grup projeleri tamamlamışlardır. Öğrenciler ayrıca, tamamladıkları projeyle ilgili olarak hem kendilerini hem de gruptaki akranlarını performans açısından değerlendirmişlerdir. Değerlendirmeler 0'la 100 arasındadır ve her öğrenci için bir akran değerlendirme notu elde etmek üzere ortalanmıştır. Bu not, hem öğrencinin kendine verdiği notu, hem de akranlarının ona verdiği notu içermektedir. ANCOVA sonuçları, akran değerlendirme notlarının, öğrencinin ara sınav notunu tahmin edebildiğini ortaya koymuştur; fakat bu sonuç kısıtlı bir hedef kitle için geçerlidir. Ayrıca, öğrencilerin sınav notları sınıftan sınıfa farklılık göstermektedir.
Anahtar Kelimeler: Grup Çalışması, Öğrenci Başarısı, Öğrenci Katılımı, Akran Değerlendirme, Değerlendirme

1. INTRODUCTION (GİRİŞ)

The education process in universities goes through a change in learning skills from individualized study skills to collaborative learning skills such as cooperative learning, team learning and problem-based learning [1 and 2]. Somervell perceives self-evaluation, peer evaluation, and collaborative assessment as promoting student-centered approaches in education [3]. Student-centered instruction and the idea of collaborative, team-based and problem-based learning lies in the core of the constructivist learning principles [4], which are the rising trend in education within the last few decades. Collaborative learning is a situation in which two or more people (a small group with 3-5 subjects, or even a class) learn or attempt to learn something (or produce a project) together [5]. Group members perform learning activities such as carrying out a project in different forms of interaction. The final outcome from the project affects each individual's grade. Individual grades are formed from the combination of individual and team performance on exams, assignments and the completed projects. The individual performance of a member can be evaluated by exams and in some way by assignments. However, it is sometimes difficult to evaluate the contributions of a team member to the team and to the project. Assessments of the team member's effort can be evaluated by peer evaluation [6].

Kane & Lawler listed the types of peer assessment as peer nomination, peer rating, and peer ranking [7]. To determine least and best performing participants, peer nomination can be used. Group members can be distinguished from each other by peer ranking. To evaluate the contribution of subjects to a project, peer rating is the best way to use. In this study, peer rating type of assessment is used.

Some educational contexts require students to work in groups to complete a project that may not as affectively be completed individually. While group work has some advantages, there are some disadvantages of working in groups, as well. Group members are supposed to share the workload. But, in some groups, workloads are not shared proportionally or group members do not contribute equally. In such a case, other group members work harder than the irresponsible peers. It is difficult to grade each team member systematically, according to their performance. One way to minimize the problems faced with this dilemma is to require students to evaluate their peers and themselves. To know being evaluated by peers causes an increase in the performance of teams [8 and 9]. Peer and self-assessments can simplify learning as the student more increasingly involves in learning and in the evaluation process [10]. However, this method may also have some contradictions. One of the problems is that a team member may evaluate more or less than he or she should. Earlier studies show that blind peer evaluation systems may decrease the inflation effects on assessment [11]. There are indications that students' performance evaluations of their peers are affected by whether they also do self-assessment [12 and 13]. This study concentrated on various courses and tried to identify whether peer evaluations are equally effective in predicting students' individual performance. Students were required to rate their peers as well as themselves.

There are several examples of use of peer evaluations in the classroom. Some earlier studies indicate that there is a high level of correlation between the self-evaluation grade and the grade given by the instructor/tutor [13 and 14]. Moreover, Fallows & Chandramohan used combination of self- and peer assessment in literature classes effectively [15]. In another study, Freeman investigated the contribution of peer assessments to group project in oral presentation

of 210 undergraduate business students by using a 22-point assessment guide [16]. Although the peer assessment marks were not significantly different from teacher marks, the correlation between them was low.

Pope investigated the impact of stress in peer evaluations [13]. For this purpose, he setup research groups where (1) only the tutor rated the performance, (2) both the tutor and the student rated the performance, (3) the tutor and student's peers rated the performance and (4) all parties rated the student's performance. Students worked in groups to complete projects. In each group students were informed about the way they would be evaluated, respective of the experimental setting they were in. He identified a significant correlation between all mentioned experimental groups. Students in self or peer evaluation groups reported more stress than the students in the tutor only groups. It was also found that stress differed between males and females as a function of the experimental conditions.

Similarly Rudy et al. compared students' self-assessments with the assessments of peers and faculty during an interviewing course in medical education [14]. Written evaluations showed that peers were more tolerant than faculty when rating their classmates. Students were most critical of their own performances, as well.

Williams et al. investigated peer evaluation as a source of feedback for students in a qualitative study [9]. Participants reported that while the feedback they received from their peers was useful, peer assessment did not directly affect their performance in the classroom. He concludes that "professors should structure peer-feedback during a project, with peer-evaluation at the end of the project" (p.698).

2. RESEARCH SIGNIFICANCE (ÇALIŞMANIN ÖNEMİ)

Falchiko & Goldfinch did a meta-analysis study by searching through six different databases and got forty eight quantitative peer assessment studies comparing peer and teacher marks [17]. The mean correlation between the teacher and peer ratings overall was found to be .69. This indicates that peer assessments of students were close to teachers' assessments of the students. 12 (25%) of the 48 studies were from Science and Engineering field, 14 (29%) were from Social Science and Art field, 19 (40%) were from Medicine and the rest were from Business field. There were only two Computer Science and one English Teacher Education fields out of the 48 studies. Considering that the study was a meta-analysis on peer evaluations, it can be said that there is a need to investigate the matter of peer assessment in the Computer Education and Instructional Technology field.

The studies above and in the literature investigate peer evaluation from different perspectives, such as stress, male and female difference, teacher and student ratings and so forth; but very few, if not at all, focus on students' individual performance.

The purpose of this study is to determine whether peer evaluations can effectively contribute to individual student course grades; in other words, whether student individual performance indicators correlate with the assessments done by peers.

3. METHODOLOGY (YÖNTEM)
3.1. Participants and Data Collection
(Çalışma Grubu ve Veri Toplama)

Participants were 390 undergraduate students from four different instructional design courses. The courses were namely, Instructional Design; Human Computer Interaction; Design, Development and Evaluation of Educational Software; and Multimedia Design and Development. This case study required students to form and join a team, decide on

projects, perform analyses, and report their findings. Membership in any particular team was voluntary, but each student had to become a member of a team. Students worked in groups of 2 – 7 to complete a project. Data were collected between spring 2008 and spring 2011, each semester from one to two different courses. Students took a midterm exam in each course. They were individually graded in this exam. Students completed their team projects near the midterm exam. They did peer and self-rating evaluations of their team members regarding the project they completed. Peer evaluations were on a scale of 0 to 100 and were averaged for each individual student to produce a mean peer evaluation score. This score included a self-score and scores given by the peers. Students were required to justify each score they gave, whether the score belonged to the student himself/herself or the teammate. The peer evaluation scores were not directly disclosed to students but were used to calculate the individual scores students received for the project. By checking their teammates' scores, students were able to evaluate their performances.

3.2. Data Analysis (Veri Analizi)

An ANVOCA analysis was run to predict the midterm score (dependent variable) where the type of course was the independent variable and peer evaluation score was the covariate to identify the peer evaluation's effect on the midterm score in connection with the type of the course. A subsequent parameter estimates were obtained to predict the midterm score and to identify the power of the peer evaluation on the midterm score.

4. RESULTS (BULGULAR)

The results show that the average midterm exam scores of the courses accumulated around sixties (see Table 1). Students assign very high scores to rate their peers' performances (an overall mean score of 87.99 out of 100); but still then, some students receive 0 (zero) for their performances, indicating that they never contribute to the team project.

Table 1. Descriptive information about the study variables
(Tablo 1. Çalışmanın değişkenleri hakkındaki betimsel sonuçlar)

	N	Midterm Exam				Peer Evaluation			
		Max	Min	Mean	SD	Max	Min	Mean	SD
Design, Development and Evaluation of Educational Software	95	95.00	20.00	61.20	14.98	100.00	65.00	89.61	7.94
Instructional Design	128	95.00	14.00	63.16	15.28	98.80	.00	84.80	13.58
Multimedia Design and Development	114	84.00	.00	59.08	12.44	100.00	80.00	90.66	4.39
Human Computer Interaction	53	87.00	26.00	56.47	15.37	97.67	55.00	87.04	10.71
Total	395	95.00	.00	60.56	14.54	100.00	.00	87.99	10.13

Logically, students' midterm results of different courses would possibly be different from each other because each course is designed to assess a different trait of students. This difference is not the main concern of this study but it is something that needs to be addressed in the analyses in order to better express the effect of the peer evaluation. Especially, it is important to understand whether there is any interaction between the course type and peer evaluation in predicting the midterm scores. In other words, it is necessary to know whether systematically the influence of one variable depends on

the level of another variable. For this reason, a preliminary one-way analysis of covariance (ANCOVA) was run with midterm exam being the dependent variable and the course type being the independent variable. The peer evaluation was considered as a covariate. A simple regression could also be used to predict the relationships between these variables but a regression analysis does not directly tell whether the independent variables correlate with each other while predicting the dependent variable.

The homogeneity-of-regression is an assumption for the ANCOVA results to be meaningful. As part of ANCOVA the homogeneity-of-regression (slopes) was tested to identify any interaction between the independent variable and the covariate. The results showed that the relationship between the peer evaluation and the midterm exam was not significantly affected by the influence of the course type, $F(3, 380)$ = 2.278, p = .079. Another assumption in ANCOVA is the homogeneity of the group variances. Levene's test of equality of error variances showed that the variance in students' midterm scores in different courses did not differ significantly, $F(3, 384)$ = 2.550, p = .055.

The ANOVA returned a significant result for the course type, $F(3, 383)$ = 3.773, p = .11, as well as the peer evaluation $F(3, 383)$ = 4.548, p = .34 (see Table 2). This means that some of the total variance in the midterm exam scores was accounted for by the course type, after controlling for the effect of the peer evaluation scores.

Table 2. Analysis of co-variance for midterm exam by course type and peer evaluation
(Tablo 2. Ders türü ve akran değerlendirmesine göre organize edilmiş ara sınav için ANOVA sonuçları)

Source	SS	df	MS	F	p
Corrected Model	2,916.50*	4	729.12	3.52	.008
Intercept	10,118.00	1	10,118.00	48.88	.000
Peer Evaluation	941.56	1	941.56	4.55	.034
Course	2,343.48	3	781.16	3.77	.011
Error	79,287.37	383	207.08		
Total	1,508,398.50	388			
Corrected Total	82,203.86	387			

* R^2 = .035 (Adjusted R^2 = .025)

Table 1 lists the mean midterm exam scores of each course. As seen in the table, the lowest mean score (56.47) belonged to the Human Computer Interaction course. The highest mean score (63.16) belonged to the Instructional Design course. The results show that at least these two scores – as being the lowest and the highest scores – are significantly different from each other.

As stated earlier, the effect of the course type was already expected. For the purposes of this study, the most important take away in this result is the impact of the peer evaluation on the complete formula. Now that this is known, the question becomes how much the peer evaluation affects the midterm exam. Therefore, the next analysis focused on the power of the factors in predicting the midterm exam scores.

Parameter estimates were collected to holistically show the power of the independent variables in predicting the midterm scores (see Table 3). In this analysis, the course type was converted into four separate dichotomous variables to be entered into the analysis. Three of the newly created variables, namely (1) Design, Development and Evaluation of Educational Software, (2) Instructional Design, and (3) Multimedia Design and Development were utilized in the estimates.

The estimates indicated a significant relationship between the dependent variable and the variables of the Peer Evaluation and Instructional Design with R = .188 and R^2 = .035 (see Table 3).

Table 3. Parameter estimates of the analysis of co-variance for midterm exam by course type and peer evaluation
(Tablo 3. Ders türü ve akran değerlendirmesine göre organize edilmiş ara sınav için ANOVA sonuçlarına ait parametre tahminleri)

Parameter	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	p
Intercept	42.67	6.77		6.31	.000
Peer Evaluation	.16	.07	.11	2.13	.034
Design, Development and Evaluation of Educational Software	4.32	2.48	.13	1.74	.082
Instructional Design	7.05	2.36	.23	2.99	.003
Multimedia Design and Development	2.29	2.41	.07	.92	.357
Human Computer Interaction	0*				

* This parameter is set to zero because it is redundant

From the regression coefficients, it is possible to say that a student who enrolled in the Instructional Design course, on average would score 7.05 points higher in midterm exam than a student enrolling in the Human Computer Interaction course. Enrolling in any other course would not bring any significant impact in the midterm exam score (in other words, students' midterm scores of the other two courses were not much different from the scores of the Human Computer Interaction course and scores of each other). Similarly, for a student to receive one point higher peer evaluation score, would mean .16 point increase in the midterm exam score. When the standardized beta coefficients are compared, it can be seen that the peer evaluation is about half times (.11/.23) as powerful as enrolling in the instructional design course.

5. CONCLUSIONS (SONUÇ)
Mainly, the literature reported in this study indicates that student performances are somewhat related with student peer evaluations. The results of this study marginally agree with the literature. The type of the course and how peers rate their classmates do determine the scores students receive from exams. It was possible to predict that students in the Instructional Design course would receive on average 7.05 points higher than students in the Human Computer Interaction course. It was also possible to predict that peer evaluation scores were almost half times as powerful in predicting the midterm exam scores as the Instructional Design course scores were. However, the findings advise us to be cautious by pointing out that the predictive power of the evaluations can be quite limited. The results yield a .035 R^2 value for the ANCOVA model. This value translates into only 3.5 percent impact on the population. And this percentage includes the impact of the type of the course in addition to the peer evaluation factor. The most of the variance in the population's exam scores are still not accounted for by the model described in this study.
Another point brought forward in the literature is the fact that students, peers and instructors rate student performances rather similarly [13]. While there was no direct instructor observation of the peers, the interesting result obtained in this study is that

students rate their peers and themselves quite generously. The mean peer evaluation score, as indicated previously, is 87.99. But, the overall performances in the midterm exam range from 56.47 to 63.16. These scores are about 25 points below the average score students use to rate each other. This does not mean – and is not against the literature – that the instructor scores and student scores do not correlate; but when it comes to predict one from the other, a regression model would be required and there seems to be a need for an intercept to balance the predicting and predictor scores. So, this is also a point to consider when incorporating the peer evaluation scores in any course.

One reason for all these results may be the nature of the assessments. One may argue that the student projects and student exams assess different aspects of student skills. Projects are more hands on, and more time consuming, and therefore require more student involvement. Paper-based exams are conducted on a limited time frame and mainly focus on the theoretical knowledge. But, because of the purposes of the course in question, these two factors have to depend on each other and should have many things in common to build a complete course assessment. So, one would still expect a connection between the two factors.

One last role of peer evaluations could be considered as a feedback agent for students to improve their performances. The fact that students completed the projects before the midterm exam, gives them the opportunity to use the feedback from the project as Williams et, al. indicated [9]. While the current study focused on identifying whether the peer evaluations and other performance indicators are in line, it would be interesting to see whether peer evaluations actually impact student individual performances. And therefore act not only as an assessment tool but also as an instructional methodology to improve instruction. Future studies can focus on this aspect especially in the Computers and Instructional Technology field based on the need identified from Falchiko & Goldfinch [17].

NOTICE (NOT)
In this study, 22-24 September 2011 in Elazig between the "(ICITS-2011) 5 International Computer and Instructional Technologies Symposium" presented as an oral presentation in.

REFERENCES (KAYNAKLAR)
1. Michaelsen, L.K. and Watson, W.E., (1993). Beyond Groups and Cooperation: Building High Performance Learning Teams, *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development,* Stillwater: New Forums Press, pp. 127-145. Retrieved August 20, 2011 from http://www.uky.edu/~drlane/beyond.html.
2. Watson, W.E., Barnir, A., and Pavur, R., (2010). Elements Influencing Peer Evaluation: An Examination of Individual Characteristics, Academic Performance, and Collaborative Processes, *Journal of Applied Social Psychology*, *40*(12), 2995–3019.
3. Somervell, H., (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education, 18*(3), 221-233.
4. Duffy, T.M. and Cunningham, D.J., (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology*. NY: Macmillan.

5. Dillenbourg, P., (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed) *Collaborative Learning: Cognitive and Computational Approaches*. (pp. 1-9) Oxford: Elsevier. Retrieved August 20, 2011 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.167.4896&rep=rep1&type=pdf,

6. Blackmore, J.A., (2005). A critical evaluation of peer review via teaching observation within higher education, *International Journal of Educational Management*, *19*(3), 218 – 232.

7. Kane, J.S. and Lawler, E.E., (1978). Methods of peer assessment. *Psychological Bulletin, 85*, 555-586.

8. Ward, M.A., (2005). An empirical assessment of peer review reliability: The ipsitive instrument paradox. *Journal of Education for Business, 81*, pp. 35-40.

9. Williams, B.C., He, B.B., Elger, D.F., and Schumacher, B.E., (2007). Peer evaluation as a motivator for improved team performance in Bio/Ag Engineering Design Classes, *International Journal of Engineering*, *23*(4), pp. 698-704.

10. Ellis, G., (2001). Looking at ourselves—self-assessment and peer assessment: Practice examples from New Zealand. *Reflective Practice*, *2*(3), 289-302.

11. Lejk, M. and Wyvill, M., (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education, 26*(6), 551–561.

12. Saavedra, R.K. and Kwun, S., (1993).Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, *78*(3), *450–462.*

13. Pope, N.K., (2005). The impact of stress in self- and peer assessment. *Assessment & Evaluation in Higher Education*, *30*(1), 51–63.

14. Rudy, D.W., Fejfar, M.C., Griffith, C.H., and Wilson, J.F., (2001) Self- and peer assessment in a first-year communication and interviewing course. *Evaluation and the Health Professions*, *24*(4), 436-445.

15. Fallows, S. and Chandramohan, B., (2001). Multiple approaches to assessment: Reflections on use of tutor, peer and self assessment. *Teaching in Higher Education 6*(2), 229–246.

16. Freeman, M., (1995) Peer assessment by groups of group work, *Assessment and Evaluation in Higher Education*, *20*, 289-300.

17. Falchiko, N. and Goldfinch, J., (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287-322.