

EURASIAN JOURNAL OF EDUCATIONAL RESEARCH

A Bimonthly Peer-Reviewed Journal, Year: 14 Issue: 57 / 2014
İki Ayda Bir Yayınlanan Hakemli Dergi, Yıl: 14 Sayı: 57 / 2014

FOUNDING EDITOR / Kurucu Editör
Veysel Sönmez, Hacettepe University, Ankara, TURKEY

EDITOR / Editörler
Şakir Çankır, Ankara University, Ankara, TURKEY

CO- EDITORS

Ayşe Çiftçi, Purdue University, IN, USA
Şenel Poyrazlı, Penn State University, PA, USA
Ramazan Baştürk, Pamukkale University / TURKEY
Ahmet Aypay, Eskişehir Osmangazi University, Eskişehir, TURKEY
Bülbin Sucuoğlu, Ankara University, Ankara, URKEY
Kürşad Yılmaz, Dumlupınar University, Kütahya, TURKEY
Hakan Atılgan, Ege University, İzmir, TURKEY

INTERNATIONAL EDITORIAL BOARD / Uluslararası Editörler Kurulu

Anita Pipere, Daugavpils University, LATVIA
Aslı Özgün Koca, Wayne State University, USA
Beatrice Adeara, West Chester University, USA
Birgit Pepin, Sor-Trondelag Univ. College / NORWAY
Gerry McNamara, Dublin City University, IRELAND
Danny Wyffels, KATHO University, Kortrijk, BELGIUM
David Bridges, Cambridge University /UK
Ekber Tomul, Mehmet Akif Ersoy University, Burdur TURKEY
Erdinç Duru, Pamukkale University, Denizli, TURKEY
Fatma Hazır Bıkmaz, Ankara University, TURKEY
Hasan Hüseyin Aksoy, Ankara University, Ankara, TURKEY
Iordanescu Eugen, Lucian Blaga University of Sibiu, ROMANIA
Joe O'Hara, Dublin City University / IRELAND
Sven Persson, Malmö University, Malmö, SWEDEN
Theo Wubbels, Univeristiy of Utrecht/ NETHERLAND
Ürsula Casanova, Arizona State University, USA
Yusif Mammadov, Azerbaijan State Pedagogy University, Baku, AZERBIJAN

SAHİBİ VE YAZI İŞLERİ MÜDÜRÜ / Publishing manager
Özer Daşcan

EDITORIAL OFFICE / Yayın Yönetim
Dilek Ertuğrul
Anı Yayıncılık, Kızılırmak Sokak 10/A
06640 Bakanlıklar Ankara, TURKEY
ejer.editor@gmail.com
Tel: +90.312 425 81 50 pbx Fax: +90.312 425 81 11

Printing Date / Basım Tarihi: 01. 11. 2014
Printing Address / Matbaa Adresi: Sözkese Mat. İ.O.S. Mat. Sit. 558 Sk. No:41 Yenimahalle-Ankara
Yayın Türü: Yaygın Süreli
Cover Design / Kapak Tasarımı: Anı Yayıncılık
Typography / Dizgi: Kezban KILIÇOĞLU
The ideas published in the journal belong to the authors.
Dergide yayınlanan yazıların tüm sorumluluğu yazarlarına aittir

Eurasian Journal of Educational Research (ISSN 1302-597X) is a quarterly peer-reviewed journal published by Anı Yayıncılık Eğitim Araştırmaları (ISSN 1302-597X) Anı Yayıncılık tarafından yılda dört kez yayınlanan hakemli bir dergidir.
© 2014 ANI Publishing. All rights reserved. © 2014 ANI Yayıncılık. Her hakkı saklıdır.

Eurasian Journal of Educational Research (EJER) is abstracted and indexed in;
Social Scisearch,
Journal Citation Reports/ Social Sciences Editon,
Higher Education Research Data Collection (HERDC),
Educational Research Abstracts (ERA),
SCOPUS database,
EBSCO Host database, and
ULAKBİM national index.

CONTENTS

İçindekiler

Crossed Random-Effect Modeling: Examining the Effects of Teacher Experience and Rubric Use in Performance Assessments

Adnan Kan, Okan Bulut..... 1-26

Graduate Students Perceptions' on Multicultural Education: A Qualitative Case Study

Hasan Aydın, Betül Tonbuloğlu..... 29-50

Opinions of Secondary School Students on the Effects of Disciplinary Regulations and Practices on the Freedom of Thought and Expression

Pelin Taşkın..... 51-72

The New Leadership Model of University Management for Innovation and Entrepreneurship

Gamze Sart..... 73-90

Investigation of Vocational Interest and Preference in Terms of Gender and Socio-economic Status

Kaan Zülfikar Deniz, Ersin Türe, Aslı Uysal, Tuba Akar ... 91-112

Measuring Essay Assessment:

Intra-rater and Inter-rater Reliability

Ulaş Kayapınar113-136

REVIEWERS of the 56th ISSUE

57. Sayı Hakemleri

Göksu Gözen

Binnur Yeşilyaprak

Adnan Erkuş

Zekeriya Nartgün

Burhan Özfidan

Tufan Adıgüzel

İlhan Kaya

Gülşah Başol

Hakan Atılğan

Funda Nayır

Nihan Demirkasımoğlu

Cevat Elma

Kürşad Yılmaz

Şenel Poyrazlı

Crossed Random-Effect Modeling: Examining the Effects of Teacher Experience and Rubric Use in Performance Assessments

Adnan KAN**

Okan BULUT***

Suggested Citation:

Kan, A., & Bulut, O. (2014). Crossed random-effect modeling: examining the effects of teacher experience and rubric use in performance assessments. *Eurasian Journal of Educational Research*, 57, 1-28. doi: [dx.doi.org/10.14689/ejer.2014.57.4](https://doi.org/10.14689/ejer.2014.57.4)

Abstract

Problem Statement: Performance assessments have emerged as an alternative method to measure what a student knows and can do. One of the shortcomings of performance assessments is the subjectivity and inconsistency of raters in scoring. A common criticism of performance assessments is the subjective nature of scoring procedures. The effectiveness of the performance assessment procedure depends highly on the quality and coordination of teacher and rubric. To gain a better understanding of the interaction between teachers and performance assessments, it is crucial to examine the effects of teacher-related factors and how teachers interact with scoring rubrics when grading performance assessments. One of these factors is teachers' work and scoring experience. When grading performance assessments, the experienced teachers may be expected to grade student performances more objectively through their experience in instruction and evaluation than the teachers with less teaching and scoring experience.

Purpose of Study: This study investigates the impact of rubric use and teaching experience on teachers' scoring behaviors in performance assessments. The effects of teaching experience and rubric use on the consistency of scores assigned by teachers is examined through an empirical study.

**Corresponding author: Dr., Department of Education Sciences, Gazi University, Ankara, Turkey. E-mail: adnankan@gazi.edu.tr

***Dr., Department of Educational Psychology, University of Alberta, Edmonton, Alberta, Canada. E-mail: bulut@ualberta.ca

Methods: Crossed random-effects modeling was used to estimate rater effects, consistency among the teachers, and the effect of teaching experience.

Findings and Results: Results indicated that lack of a scoring guide may cause the teachers to establish their performance criteria and score tasks inconsistently. When teachers used a rubric, inter-rater reliability substantially increased. Experienced teachers and teachers with little teaching experience exhibited different severity patterns in scoring.

Conclusions and Recommendations: Based upon the results of this study, it appears that teachers who have more teaching experience tend to score performance tasks more leniently than teachers who do not have long years of teaching experience. The differences in the teachers' scoring due to their teaching experience became negligible when all teachers used a scoring rubric. In addition to teaching experience, the potential effects of other external factors should also be considered to make the use of rubrics more effective in performance assessments. This study illustrated an alternative methodology to estimate variance components and the effects of fixed factors within the same analysis. A big advantage of this modeling approach over generalizability theory is that it allows for the separation of random and fixed effects from each other. Although the findings of this study enrich the limited knowledge about the effects of rubric use and teaching experience on teachers' scoring behaviors, further research is needed to understand the reasons why these factors are influential.

Keywords: Performance assessment, rubric, teaching experience, reliability, rater effects, crossed random effects model.

In the last two decades, most educators in K-12 and higher education institutions have started focusing on critical thinking and problem solving abilities rather than factual knowledge and lower-level cognitive skills. To prioritize critical thinking and problem solving skills in instruction, the evaluation policies and procedures have also been changing from the conventional testing of knowledge to —evaluation for learning! (Dochy, Gijbels, & Segers, 2006). To measure critical thinking, traditional methods of assessment (e.g., paper and pencil assessments, multiple-choice tests) do not seem to be adequate. Standardized test scores and traditional ways of reporting grades provide a narrow and arbitrary measuring system that does not give any additional information about instructional purposes (Routman, 1991). Therefore, the inadequacy of the existing assessment methods has led to the development of alternative testing methods—such as performance assessments—that prompt students to use higher-order thinking skills such as analysis, synthesis, and evaluation.

Although standardized tests have dominated student assessment systems and assessment policies such as No Child Left Behind for years, educators in the United States and other nations have been able to move to utilizing performance

assessments and portfolios as an alternative against standardized testing for assessing student performances. The essence of reality in an assessment is to provide more valid information about the competence of the student and to reflect complexity in the real world rather than solely focusing on the truth where the context is the only source for the learning process (Darling-Hammond & Synder, 2000). As Palm (2008) mentioned, the performance assessment method is viewed as providing more opportunities to measure complex skills and communication, which are considered important competencies and disciplinary knowledge needed in today's society. In comparison to conventional assessment methods, performance assessments allow students to become more open in their responses (Messick, 1996). On such assessments, students are required to perform a task rather than select an answer from a ready-made list, such as multiple-choice items. As a result, performance assessments are able to take control of more elusive part of learning in such a way that students are required to deal with realistic and authentic problems.

Despite their many advantages over traditional assessment methods, performance assessments have not been considered as the main tool for student assessment. A common criticism of performance assessments is the subjective nature of scoring procedures. While questions in traditional assessments can easily be scored as right or wrong, the difference is not as clear-cut with performance assessments (Brualdi, 1998). Two important steps are required to minimize the role of subjectivity in performance assessments. First, a scoring mechanism should be developed to grade performance assessments in a reliable way. This scoring mechanism should consist of performance level descriptors (PLDs) that provide information to teachers and students about the skill and knowledge a student needs to demonstrate along with a scoring rubric that indicates what criteria should be used to evaluate students' performances. The second step is to provide training to teachers or raters about how to use PLDs and rubrics to make judgments about students' performance on a task.

The effectiveness of the performance assessment procedure highly depends on the quality and coordination of teacher and rubric. To have a better understanding of the interaction between teachers and rubrics, potential factors that may influence scoring procedures should be carefully examined. One of these factors is teachers' work and scoring experience. Experienced teachers may be expected to grade students' performances more objectively because of their experience in instruction and evaluation. Of course, this does not necessarily mean that teachers with little experience would score inconsistently. Rather, teachers who have recently started their careers can be more objective and consistent in scoring performance assessments since performance-based assessment systems have recently been a core part of teacher education programs. To gain a better understanding of the interaction between teachers and performance assessments, it is crucial to examine the effects of teacher-related factors (e.g., teaching experience) and how teachers interact with scoring rubrics when grading performance assessments. This study investigates the impact of teaching experience and rubric use on teachers' scoring behaviors in

performance assessments. The effects of teaching experience and rubric use on the consistency of scores assigned by teachers is examined through an empirical study.

Rubrics in Performance Assessments

Performance assessments require grading strategies that are commonly used in the applied sciences, performing arts, fine arts, and Olympic competitions. For example, in the context of the science laboratory, students are graded based on their performance of manipulating variables, using scientific apparatus, identifying hypotheses, making measurements and calculations, organizing and managing data, and the communication of results (Slater & Ryan, 1993). Studies that have closely looked at performance assessments indicate that, if the evaluation criteria are clear and there are available examples to show levels of competency, performance assessments are highly consistent across different raters (Kulm & Malcom, 1991; O'Neil, 1992). To assure that raters implement a consistent grading procedure across all examinees, scoring guidelines called —rubrics are used for scoring performance assessments.

Rubrics are the vehicles that provide a useful mechanism to translate students' achievement into assessments (Schafer, Swanson, Bené, & Newberry, 2001). Rubrics provide a description of various levels of performance for a certain task, and define what varying levels of mastery should look like (Hafner & Hafner, 2003). To produce valid and reliable results, a rubric should provide enough information to help raters to assess student performances (Stuhlmann, Daniel, Dellinger, Denny, & Powers, 1999). Rubrics usually consist of a scoring scale from 3 to 5 points depending on the evaluation criteria. Each of the points corresponds to a certain level of performance on a particular task (Wiener & Cohen, 1997). The scoring criteria on the rubric must correspond with the pre-specified standards, and distinguish between levels of performance (Farr & Tone, 1998).

Previous research has indicated that when raters use rating scales or rubrics inconsistently, it may result in a wide range of misleading scores (Hadden, 1991; Wigglesworth, 1994). Inconsistent use of rubrics may occur due to a lack of understanding of the construct or the rubric. Raters may attempt to use wide range of non-criterion information when scoring performances (Ang-Aw & Goh, 2011). In classrooms, teachers can easily familiarize themselves with scoring rubrics because they are supposed to know the construct to be measured and performance levels to be achieved. However, it does not guarantee that each teacher comprehends a rubric in the same way. As teachers become more competent and experienced with instructional practices, they can use scoring rubrics and performance assessments more effectively in their classrooms.

The teacher's role in assessments is crucial because they make judgments about the quality of a student's performance. Brennan (2000) contends that, in most performance assessments, raters do not often cause a considerable variability in observed scores. However, this does not necessarily mean that this variability is always negligible, and so it should be ignored. Differentiation among raters' scoring behaviors may be an important source of score variation because raters' scoring

behaviors have a direct impact on the reliability and validity of performance assessments. There are several factors that may affect raters' scoring performances, such as rater training, rubric use, quality of rubrics, and multiple scoring occasions. In addition, teacher characteristics such as teaching background, teaching experience, and scoring experience may also influence the way teachers approach the scoring of performance assessments.

Receiving decent training on how to score performance tasks using a rubric may positively influence the way that raters understand and score tasks in an assessment (Schafer et al., 2001; Stuhlmann et al., 1999). Of course, training raters does not necessarily assure that all raters interact with the scoring rubric and the assessment in the same way. Despite receiving the same training about scoring rubrics, raters may evaluate student performances differently because of their own construct of a good performance (Lumley, 1998; Schafer et al., 2001). As Eckes (2008) highlighted, raters may differ not only in the way they understand and apply the scoring criteria, but also in the level of subjectivity they incorporate into scoring, and in the degree to which their ratings are consistent across examinees, scoring criteria, and tasks. Even though teachers use the same rubric, they may differ in terms of their scoring severity based upon their familiarity with rubrics and performance assessments.

Factors such as the background and scoring experience of the raters have also been shown to affect raters' scoring practices (Eckes, 2008; Hamp-Lyons, 1991). In addition to scoring experience, the teaching experience of teachers is also an important criterion that has been used by many countries (e.g., England, France, and Canada) to recruit raters who score the nation-wide exams. Teaching experience as a rater selection criterion has frequently become embedded in the cultural expectations associated with the tests (Royal-Dawson & Baird, 2009). For instance, in the UK, the Assessment and Qualifications Alliance (AQA) requires raters selected for national examinations to have suitable academic qualifications, and at least three terms of teaching experience that should be recent and relevant.

Although teaching experience has been a criterion for rater selection, there is very little research about the effect of teaching experience on scoring and how to select raters who can score more accurately (e.g. Pinot de Moira, 2003; Royal-Dawson & Baird, 2009). Since performance and portfolio assessments are relatively newer techniques in comparison to conventional testing, new teachers may be expected to be more familiar with performance assessments and rubrics for scoring portfolios and performance tasks. However, experienced teachers are also advantageous because they are more likely to have a better sense of assigning coherent scores across examinees based upon their experience. Previous research suggests that raters with little experience tend to score tasks more severely than experienced raters (Shohamy, Gordon, & Kraemer, 1992; Weigle, 1999). However, Myford and Mislevy (1994) and Meyer (2000) found that teaching experience had only a negligible impact on predictions of rater severity.

These rater-related issues bring us to the question of whether severity in scoring or inconsistency among raters is more noteworthy. Since performance and portfolio

assessments in classrooms are mostly graded by a single teacher, the issue of inconsistency among raters is not likely to occur. In that case, exposing all of the students' responses to the bias of a single rater, which is known as rater severity, is a more serious issue. However, in large-scale testing programs, the main concern is often inconsistency rather than severity in scoring. Variations across raters in scoring severity can be accounted for by adjusting candidates' scores (Baird & Mac, 1999). To examine the consistency among raters or across scoring occasions, different measures of consistency such as inter-rater and intra-rater reliability coefficients can be used.

Reliability of Performance Assessments

Raters, tasks, and other sources of score variation can influence the precision of performance assessments (Brennan, 2000). Each source brings different types of issues that may influence the consistency of scoring within or between raters. There are several methods to determine consistency of scores across raters or occasions, such as Cohen's kappa statistic (Cohen, 1960), Pearson correlation coefficient, and coefficient alpha (Crocker & Algina, 1986). All of these reliability coefficients have been used to examine a single source of measurement error in students' scores. An alternative way to examine the consistency of scores is to obtain measurement estimates based on a summary score for each participant, taking into account the extent to which each judge influences the score. The most common example of this method is generalizability theory, which allows for estimating reliability by examining multiple sources of errors and their possible interactions simultaneously (Brennan, 2001; Shavelson & Webb, 1991).

Unlike a single error term in classical test theory (CTT) that is the total variation in scores due to error, generalizability (G) theory allows for the examination of multiple sources of error simultaneously (Shavelson & Webb, 1991). G theory extends the idea of total variation consisting of true score and error variations in CTT by ascribing variations in observations to specific sources, such as persons, raters, tasks, etc. In G theory, sources of random variation are defined as facets. A generalizability model can separate out random variations due to each facet that contribute to the total error variation. Detecting the variation due to each facet can provide a mechanism for optimizing the reliability of performance assessments.

In generalizability studies, participants, tasks, and raters are either nested or crossed. Crossed designs are those where every condition of each facet is repeated for every condition of the other facets (e.g., each rater grades all tasks responded to by all persons). Differences in rater severity are taken into account at the level of the individual person, facets, or group (Stemler, 2004). For a study with three crossed random facets (e.g., person x rater x task), random variation due to each of these facets, and their two-way and three-way interactions, are estimated. In this example, the total observed score variance (X_{prt}) is decomposed into seven variance components:

$$\sigma^2(X_{prt}) = \sigma_p^2 + \sigma_r^2 + \sigma_t^2 + \sigma_{pr}^2 + \sigma_{pt}^2 + \sigma_{rt}^2 + \sigma_{prt}^2 \quad (1)$$

Variance components are used to compute a generalizability coefficient that is the ratio of score variance to the sum of score variance and residual variance. The generalizability coefficient is analogous to the reliability coefficient in CTT (Brennan, 2000; Shavelson & Webb, 2005). As intraclass correlation coefficients, the generalizability coefficient can be computed by dividing the score variance among persons by the sum of score variance and error variance. In a crossed person-by-item design, the generalizability coefficient can be shown as follows:

$$E\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)} \quad (2)$$

where $\sigma^2(p)$ is universe score variance or true score variance and $\sigma^2(\delta)$ is the relative error variance of the scores, which is basically the sum of $\sigma^2(pt)$, $\sigma^2(pr)$, and $\sigma^2(prt)$. The relative error variance in G theory corresponds to residual or error variance in CTT. As with most reliability indices, the larger $E\rho^2$, the higher reliability the scores have.

Although G theory is able to break down different sources of error variation, it does not allow for the estimation of fixed covariates (i.e., fixed effects) in addition to error variation (i.e., random effects). A more sophisticated approach, mixed-effects modeling, can overcome this limitation by estimating random and fixed effects together and allowing additional covariates to be included in the model.

Random-Effects Modeling

G theory is accepted as the equivalent of the random effects model of the analysis of variance (ANOVA). The variability in scores is examined rather than estimation of universe scores because the facets are assumed to be randomly selected from some relevant universe, as in random-effects ANOVA (Ward, 1986). Facets in the generalizability model are defined as random effects in the random-effects ANOVA model. Mixed-effects modeling is a general framework that combines both linear regression and random effects modeling. Mixed-effects models are primarily used to describe linear or non-linear relationships between a dependent variable and some covariates in data that are grouped according to one or more classification factors (Pinheiro & Bates, 2000, p. 57).

Random effects in mixed-effect models correspond to facets in G theory. Random variation due to the facets and their interactions are captured by estimating random effects. Unlike hierarchical or multilevel models in which random effects must be assumed to be nested, mixed effects models allow random effects to be crossed. Because all facets are crossed with each other, this model can be called a crossed random-effects model. A crossed random-effects model with three crossed facets (e.g., persons, tasks, and raters) can be shown as follows:

$$Y_{jik} = X_{jik}\beta + T_i t_i + P_j p_j + R_k r_k + \varepsilon_{jik} \quad (3)$$

In Equation 3, Y_{jik} is the score of person j on task i determined by rater k , and β is the fixed-effect coefficient that is an identical intercept term for all the facets in the model. The next term, X_{jik} , is a design matrix that includes the fixed-effect regressors for person j on task i graded by rater k . Similar to X_{jik} , T_i , P_j , and R_k are the design matrices that represent the random-effect regressors for persons, tasks, and raters. The multipliers of the design matrices are the random-effect coefficients for persons, tasks, and raters. These random effects are assumed to have a multivariate normal distribution, $t_i, p_j, r_k \sim N(0, \sigma^2\Sigma)$, which allows each random effect to vary by its group. These terms are thought of as random variables rather than as parameters. Therefore, they are similar in this respect to the errors (Fox, 2002). The last term in the equation, ε_{jik} , is the error term for the score of person j on task i given by rater k . The error terms are also assumed to be multivariately normally distributed, $\varepsilon_{jik} \sim N(0, \sigma^2I)$. The error term, ε_{jik} , in the crossed random-effects model corresponds to the three-way interaction, $\sigma^2(\text{prt})$, in G theory.

Two-way and three-way interactions of persons, tasks, and raters can be used as sources of random variation in the crossed random-effects model. For instance, the interaction between tasks and raters (i.e. TxR) can be included in the model as a random effect to explain the random variation in the scores due to this interaction. It should also be noted that Equation 1 does not include any predictors (i.e. fixed effects) in the model. As described earlier, a crossed random-effects model can simultaneously estimate fixed effects and crossed random effects. Assuming there are fixed effects to be estimated in the model, Equation 3 can be rewritten in a short form as:

$$Y_{ijk} = X_{jik}\beta + Z_{jik}b + \varepsilon_{jik}, \quad (4)$$

where β includes all fixed-effect coefficients and b includes all random-effect coefficients in the model. The design matrices for fixed and random effects are represented by X_{jik} and Z_{jik} , and ε_{jik} is the vector of residual error of the scores across all facets.

In a crossed random-effects model, additional predictors that denote fixed effects can be either continuous or categorical. If the predictor is categorical, one of its categories becomes the reference group as a result of the dummy coding scheme. In the context of performance assessments, the joint analysis of random and fixed effects allows testing whether persons, tasks, raters, and the interactions of these components contribute to scores independently. Furthermore, the impact of predictors that are not expected to vary randomly (i.e. fixed effects) can be tested. Models with different random and fixed effects can be compared based on log likelihood and deviance statistics (Raudenbush & Bryk 2002, p. 60-61).

Interrater and Intrarater Reliability

As explained earlier, variances of random facets can be used to compute a reliability coefficient that is the proportion of true score variation to the observed

score variation. This type of reliability coefficient indicates the extent which a given assessment measures persons' true scores. Other reliability coefficients can also be computed to examine how raters influence the variability in persons' observed scores. Depending on the design of the study, two types of reliability coefficient can be used. These are inter-rater and intra-rater reliability coefficients. When multiple raters grade students' responses to the same task or different tasks, the correlation between the scores assigned by the raters becomes an indicator of inter-rater reliability (Brennan, 2003). The inter-rater reliability coefficient shows the consistency of measurements across raters and the extent to which the raters are interchangeable (Eliasziw, Young, Woodbury, & Fryday-Field, 1994). In a person \times task \times rater (i.e., $p \times t \times r$) design, the inter-rater reliability coefficient for a single occasion is:

$$\rho_{\text{interrater}}^2 = \frac{\sigma^2(p) + \sigma^2(pt)}{\sigma^2(p) + \sigma^2(pt) + \sigma^2(pr) + \sigma^2(ptr)} \quad (5)$$

When Equation 5 is reorganized using the notation in crossed random-effects modeling (see Equation 3), it can be written as:

$$\rho_{\text{interrater}}^2 = \frac{p_j + pt_{ji}}{p_j + pt_{ji} + pr_{jk} + \varepsilon_{jik}} \quad (6)$$

Inter-rater reliability is useful for measurements that are carried out on a single occasion. When raters grade items on multiple occasions, a rater's consistency (assigning the same or similar scores across occasions ($o_m = 1, \dots, M$)) can be checked by using an intra-rater reliability coefficient. For a single item in a person \times rater \times occasion design, intra-rater reliability can be defined as:

$$\rho_{\text{intrarater}}^2 = \frac{\sigma^2(p) + \sigma^2(pr)}{\sigma^2(p) + \sigma^2(pr) + \sigma^2(po) + \sigma^2(pro)} \quad (7)$$

Equation 7 can also be written again using random-effect components from the crossed random-effects model as follows:

$$\rho_{\text{intrarater}}^2 = \frac{p_j + pr_{jk}}{p_j + pr_{jk} + po_{jm} + \varepsilon_{jmk}} \quad (8)$$

The intra-rater reliability coefficient can be computed for each item separately by averaging the scores from each task over raters. As Brennan (2003) suggested, additional complexities may arise when interpreting these reliability coefficients because it may be difficult to determine whether a facet should be defined as random

or fixed in the study. Yet, they can still be very informative when researchers try to assure that raters are functioning as expected.

Method

Research Design

The aim of this study is to examine the effects of teaching experience and rubric use on teachers' grading behaviors across different occasions in a performance assessment. A performance task was given to the eighth-grade students, and their responses were graded by 17 teachers with and without a rubric. A crossed random-effects modeling approach was used to examine the individual contributions of students, tasks, teachers, and the interactions of these components with the variations in students' scores. In addition, rubric use, teachers' work experience, and the duration between grading periods are used as fixed-effects in the model. Two research questions addressed in this study are: 1) How does the rubric use influence the scoring of performance assessments? 2) Is there any interaction between teachers' work experience and their grading behaviors? The following sections describe the sample, data collection procedure, and statistical techniques used for data analysis.

Participants and Data Collection

The participants of this study were 50 eighth-grade students and 17 math teachers. The sample of teachers was chosen to represent a wide range of experience. Years of experience as a teacher ranged from one year to 26 years in the sample. Students, teachers, and performance tasks were fully crossed facets. All students in the sample responded to the same questions in a performance assessment. The responses of all the students were graded by each of the seventeen teachers.

The content of the performance assessment in this study was graphical comprehension in mathematics. In the literature, researchers have focused on the three sublevels of graph comprehension (Friel, Bright, & Curcio, 2001; Wainer, 1992). These sublevels are translation, interpretation, and extrapolation/interpolation. Translation represents a change in the form of a communication (Friel et al., 2001). Translation occurs when one interprets a graph at a descriptive level and comments on the structure of the graph (Jolliffe, 1991; Wood, 1968). Interpretation means organizing the information in the graph and determining which information is more important or necessary (Wood, 1968). Interpretation happens when one looks for relationships among specifiers in a graph or between a specifier and a labeled axis (Friel et al., 2001). Extrapolation/interpolation means understanding the data in a graph and making inferences or conclusions. Extrapolation and interpolation occur when one extrapolates or interpolates by perceiving trends shown in graphical data or by specifying implications (Wood, 1968).

By using the three sublevels of graphical comprehension, a performance assessment task was developed by the researchers along with the math teachers to assess students' graphical interpretation skills. The assessment consisted of four

open-ended questions that measure the understanding and interpretation of graphical representations of mathematical data. The first two questions included bar charts, and the other two questions were based on line charts. The first bar and line chart questions focused on the relationship between two facts for a single variable (e.g. distance by time for a bike) whereas the charts in the second bar and line chart items were about the relationship between two facts for two separate variables (e.g. phone bills by month for home and cell phones). Figure 1 shows an example of these items. All of the items required students to understand the graphical representation and create a story based on their interpretation of the graph. Students' graphical interpretation skills were evaluated based on how they interpreted the data in the graph, made conclusions, and related the information within a story.

After the students completed the performance assessment, each of the 17 teachers was asked to grade students' responses to the four questions. All questions carried an equal value in scoring. Each question was worth twenty-five points. The teachers graded the questions on four occasions. First, the teachers were asked to grade students' responses based on their own scoring criteria without a scoring rubric. Three weeks after the first grading occasion, the teachers graded the questions again, but this time they were given an analytic rubric based on the performance criteria determined for this assessment. An analytic rubric (see Appendix) was preferred because extra details in the analytic rubric were anticipated to help teachers to follow the same performance criteria.

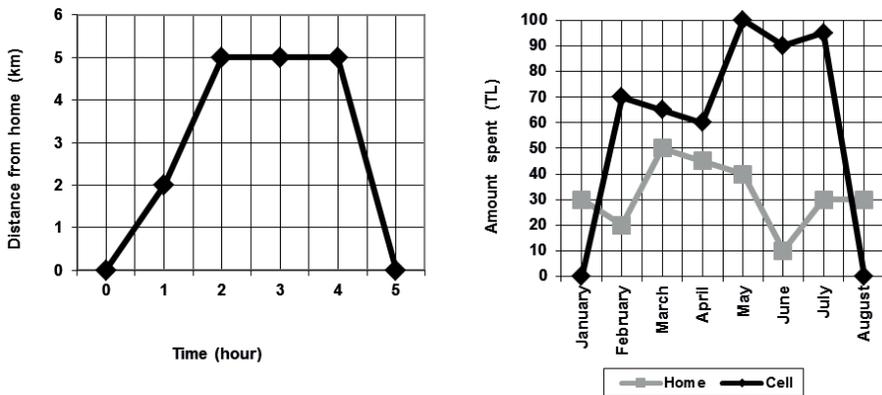


Figure 1. Example questions with line graphs for a single variable (left) and two separate variables (right) in the performance assessment.

To examine the effect of time on the consistency of raters' scores, the third occasion took place ten weeks after the second occasion. As in the first occasion, the teachers graded students' responses without using a rubric. The last grading occasion was three weeks after the third occasion. The teachers scored the questions by using the same scoring rubric that they used on the second scoring occasion. To

prevent the teachers from remembering students' responses from previous scoring occasions, the order of the students' answer sheets were scrambled, and different identification numbers were assigned to the students for each occasion.

Data Analysis

The design of this study consisted of three random facets that were fully crossed with each other. These facets were persons, tasks (i.e. questions), and raters (i.e. teachers). In addition to these facets, occasion was another source of variation, which was crossed with tasks and raters. The impact of these sources of variation on students' scores was examined using crossed random-effects models. All facets and their interactions were treated as random effects in the models. In addition to the random effects, there were two predictors that were used as fixed effects. These predictors were teachers' work experience and rubric use. Work experience (ranging from 1 to 26) represents how many years a teacher has served in a school. This was self-reported information provided by the participating teachers of the study. Rubric use was a categorical variable. Occasions in which the teachers used a scoring rubric were coded as one, and other occasions without rubric use were coded as zero.

Four crossed random-effects models were used to address the research questions explained earlier. The first two models aimed to identify variations in the scores due to the raters within occasions and across occasions. Model 1 was based on a $p \times t \times r$ design, which included random effects for persons, raters, tasks, and their two-way interactions. Model 1 was run for each occasion separately, and the inter-rater reliability coefficient was computed for each occasion by using Equation 6. Model 2 used a $p \times r \times o$ design. It estimates random effects for persons, raters, occasions, and their interactions. Model 2 was run for each task separately, and intra-rater reliability coefficients were computed. When intra-rater reliability coefficients were computed, the person \times occasion interaction was not included because the same responses from the students were graded by the raters across four occasions, and so there was no interaction between students and occasions.

Model 3 was an overall model that treated all of the facets and their interactions (i.e., persons, tasks, and occasions) as random effects. Because students' responses to the questions were fixed across occasions, the interaction between students and occasions was not included in the model. The last model, Model 4, estimated fixed effects for teachers' work experience and rubric use in addition to all random effects in Model 3. The purpose of this model was to estimate the effects of rubric use and work experience on teachers' grading behaviors after accounting for the effects of all facets.

All of the models were estimated using the *lme4* package (Bates, Maechler, & Bolker, 2011) in R (R development core team, 2012). The *lme4* package can estimate mixed-effects models with a fast algorithm for parameter estimation (see Bates, 2005; West, Welch, & Gaflechki, 2007). The *lme4* package requires a long data format (i.e. multiple rows for each person). The data structure is presented in Table 1.

Table 1*Structure of the Facets in the Data.*

Student	Question	Score	Teacher	Experience	Occasion	Rubric
1	1	15	1	5	1	0
1	2	25	1	5	1	0
1	3	20	1	5	1	0
1	4	15	1	5	1	0
1	1	10	1	5	2	1
1	2	11	1	5	2	1
1	3	8	1	5	2	1
1	4	11	1	5	2	1
1	1	20	1	5	3	0
1	2	20	1	5	3	0
1	3	20	1	5	3	0
1	4	20	1	5	3	0
1	1	13	1	5	4	1
1	2	12	1	5	4	1
1	3	9	1	5	4	1
1	4	11	1	5	4	1

Note: Only scores of student 1 given by teacher 1 is shown here.

Results

Table 2 shows that the $p \times t \times r$ design (Model 1) estimated seven variance components associated with persons, tasks, raters, person-by-task interaction, person-by-rater interaction, task-by-rater interaction, and person-by-task-by-rater interaction. The $p \times t \times r$ design was estimated separately for each of the four scoring occasions. In the first and third occasions, the largest variance component was attributable to student-by-task-by-rater interaction and students, respectively. The largest variance component was attributable to student-by-task interaction on the second and fourth occasions, accounting for more than 25% of total score variance. This was a desirable result because it indicates that most score variability was explained by the differences in students' graphical comprehension abilities rather than the raters.

A slight difference was found between the occasions in terms of inter-rater reliability estimates. Inter-rater reliability was fairly higher when the teachers were given a scoring rubric for the second and fourth occasion. This means that the scores assigned by the teachers exhibited better consistency when they were given a scoring rubric. Also, the amount of total variance decreased by as much as 70% when a scoring rubric was used for grading. The proportions of person-by-rater and rater-by-task interactions to the total variance were almost the same across four occasions, although the actual amount of these interactions was much smaller on the second and last occasions. This suggests that measurement errors due to the rater facet were reduced when a scoring guide was introduced to the raters.

Table 2

Estimated Random Effects and Inter-rater Reliability Coefficients in Model 1 Across Four Occasions.

Facet	Occasion 1		Occasion 2		Occasion 3		Occasion 4	
	<i>b</i>	%	<i>b</i>	%	<i>b</i>	%	<i>b</i>	%
p	13.65	24	3.55	23	13.73	27	3.47	18
t	1.56	3	0.28	2	1.64	3	0.49	3
r	4.19	7	2.21	14	5.26	10	3.43	20
p × t	14.55	25	4.76	31	13.69	27	4.61	27
p × r	3.86	7	0.76	5	3.72	7	0.91	5
r × t	1.24	2	0.39	3	0.99	2	0.47	3
p × t × r	18.51	32	3.57	23	14.72	25	3.52	23
ρ^2	.56		.66		.59		.65	

Note: p: person; t: question; r: rater. ρ^2 = Inter-rater reliability; *b* = estimated random effect; % = Percent of total variability.

Table 3 shows the results from the $p \times r \times o$ design (Model 2), which estimated six variance components. As explained earlier, this model was run for each question separately to compute intra-rater reliability across four occasions. The results indicated that the largest variance component was attributable to students in all four questions. Person-by-rater-by-occasion interaction was the second largest source of the variation within the questions. The highest variation (52%) in the scores due to the students was observed in the last question (question 4). This question also exhibited the highest intra-rater reliability. The rater effect was similar across the four questions. This indicates that teachers' severity or leniency in scoring followed a similar pattern across the questions.

On the second and last occasions, where the teachers used a scoring rubric, the mean scores for the questions became smaller (see Figure 2). This suggests that teachers tend to score more severely when they use a rubric for scoring the performance assessment items. In addition to the shrinkage in the mean scores, the questions showed fewer scoring errors when the teachers used a scoring rubric on the second and last occasions.

Table 3

Estimated Random Effects and Intrarater Reliability Coefficients in Model 2 Across Four Questions.

Facet	Question 1		Question 2		Question 3		Question 4	
	<i>b</i>	%	<i>b</i>	%	<i>b</i>	%	<i>b</i>	%
p	11.59	33	14.11	41	16.49	39	22.28	52
r	1.25	3	2.01	6	2.87	7	2.36	6
o	5.82	16	3.71	11	5.69	13	2.42	6
p × r	2.91	8	1.33	4	1.63	4	2.34	5
r × o	1.68	5	2.59	7	3.34	8	1.93	5
p × r × o	12.34	35	10.84	31	12.55	29	11.14	26
ρ^2	.54		.59		.59		.69	

Note: p: person; r: rater; o: occasion; ρ^2 = Intra-rater reliability; *b* = estimated random effect; % = Percent of total variability.

Table 4 contains the variance component estimates and percentages of total variation for all facets and two-way interaction effects for Model 3 and 4. Model 3 was a *p × t × r × o* design that included ten variance components. It estimated random effects for persons, raters, tasks, occasions, and their interactions. Among the ten variance components, person and person-by-task effects were the highest. The percentage of variation among the raters was high, indicating a questionable level of intra-rater reliability. Variation among the raters was almost the same as the variation due to the interaction between raters and occasions. This variation also suggests considerable discrepancies between the raters.

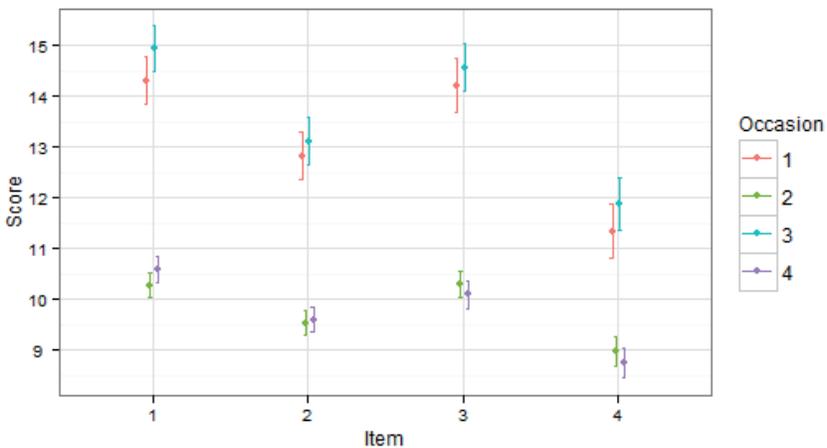


Figure 2. Mean and confidence intervals of the questions across four occasions.

Table 4*Estimated Random and Fixed Effects for Model 3 and 4.*

Effects		Model 3		Model 4	
		<i>b</i>	<i>SE</i>	<i>b</i>	<i>SE</i>
Random	p	7.58	1.36	7.58	1.35
	r	1.81	0.74	0.35	0.48
	t	0.91	0.56	0.91	0.49
	o	4.31	0.68	-	-
	p × r	1.17	0.76	1.17	0.76
	p × t	8.54	1.39	8.54	1.39
	r × t	0.29	0.41	0.29	0.41
	r × o	1.96	0.78	1.91	0.62
	r × t × o	0.58	0.54	0.54	0.54
	p × r × t	0.88	0.85	0.88	0.85
	Fixed	Rubric use	-	-	-3.63*
Experience		-	-	0.17*	0.03

(*) Significant at $\alpha = .01$

Figure 3 shows estimated random effects and their confidence intervals for each question and rater in Model 3. Question 4 seems to have a smaller random effect on average than the other questions. Although random effects for most raters were similar, there were a few raters (i.e., raters 2, 4, and 9) whose estimated random effects were relatively higher than the rest. This suggests that the scoring of some of the raters differed significantly even after all sources of potential variation were taken into account.

The last model (Model 4) estimated two fixed effects in addition to the variance components in Model 3. However, this model did not include occasion as a variance component because a fixed effect for rubric use was estimated instead of a variance component for four scoring occasions. As expected, the estimated variance components in Model 4 were very similar to the ones from Model 3, except the rater effect, which became smaller in Model 4. The results indicate that both rubric use and teachers' work experience are significant predictors of the variation in the scores. The estimated fixed effect for rubric use was -3.63, meaning that teachers scored 3.63 points less across four items when they used a rubric for scoring the items. This finding is also supported by the higher inter-rater reliability coefficients obtained from the second and fourth occasions where the teachers used a scoring rubric. It implies that teachers tend to score more severely and consistently when a scoring guide is provided.

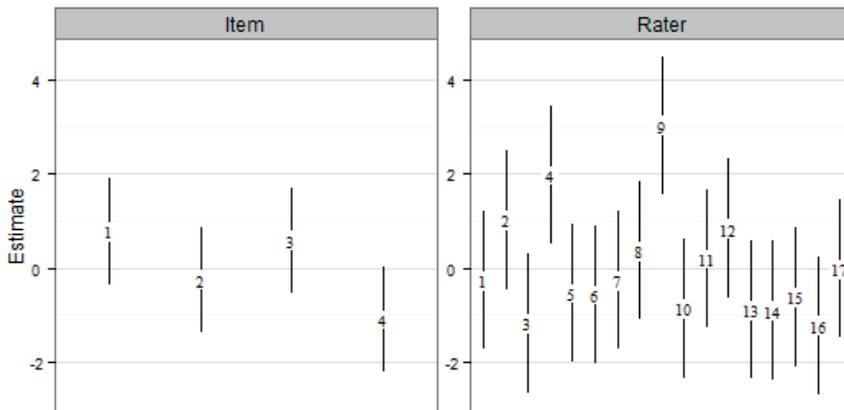


Figure 3. Random effects and confidence interval for questions and raters in Model 3.

Teachers' work experience was positively related to the way the teachers scored the performance tasks. Results indicated that the teachers with more experience assigned higher scores to the items than the teachers with less experience. The reason for this finding might be that the more experienced teachers might not be following the rubric in the same way that the teachers who have recently begun their careers do. Because new teachers are supposed to be more familiar with the nature of performance assessments, they may tend to score more strictly and consistently when grading performance assessments with or without a scoring rubric. It should be noted that although the difference in scores assigned by the experienced teachers and teachers with less experience is obvious, this difference seems to diminish when all of the teachers used a rubric for scoring the items on the performance assessment.

Discussion and Conclusion

This study utilized a random-effects modeling approach to examine whether rubric use and teachers' work experience influence teachers' scoring behaviors in a performance assessment, with particular attention paid to inter-rater and intra-rater reliabilities and changes in raters' scoring due to rubric use and teaching experience between the occasions. A mathematics assessment that consisted of four questions about graphical comprehension was administered to fifty eighth-grade students. Seventeen mathematics teachers scored the students' responses on four occasions. During the first and third occasions, teachers graded the items without a scoring guide, whereas on the second and last occasions, they scored the responses using a scoring rubric.

The results of this empirical study support the necessity of using a scoring rubric for grading performance assessments. During the first and third occasions, the lack of a scoring guide caused the teachers to establish their own scoring criteria to assess students' performances. Therefore, as an indicator of inconsistency among the

teachers, the inter-rater reliability coefficient was fairly low. However, when the teachers were given a scoring rubric on the second and last occasions, the consistency among the scores given by the teachers increased. Although the percentage of rater effects seemed to increase on the second and last occasions, the actual amount of variation components for raters became smaller.

The consistency of the scores assigned by the teachers differed across the questions. In the assessment, questions 1 and 3 were less complex compared to questions 2 and 4. Although the same evaluation criteria were applied to the questions, questions 1 and 3 could have been scored more consistently because these graphical questions asked about the relationship between two variables based on a single fact. Therefore, students' responses to them were not expected to vary much. Unlike these questions, questions 2 and 4 focused on the same type of relationship based on two facts, and they were more difficult to interpret. However, intra-rater reliability coefficients indicated that the teachers scored question 4 very consistently but they assigned very different scores for question 1 across the four occasions. The teachers took a lot of points off when they used the rubric to score question 1; they scored the same question more leniently without the rubric. This suggests that the complexity levels of tasks may also influence the way teachers interact with the rubric.

Unlike Myford and Mislevy's (1994) and Meyer's (2000) findings, which suggested that teaching experience only has a negligible effect on raters' scoring behaviors, this study indicated that teaching experience may have a considerable effect on the way teachers approach performance assessments. Based upon the results of this study, it appears that teachers who have more teaching experience tend to score performance tasks more leniently than teachers who do not have long years of teaching experience. Because this was an empirical study, it was not possible to identify whether teaching experience had a positive or negative impact on teachers' scoring behaviors. The differences in the teachers' scoring due to their teaching experience became negligible when all teachers used a scoring rubric. Providing training about the use of scoring rubrics to teachers may also help to reduce the effect of teaching experience on teachers' interaction with scoring rubrics. As Schafer et al. (2001) noted, teachers' understanding of rubrics may change based on additional factors, such as the subject and item types in an assessment. Therefore, in addition to teaching experience, the potential effects of other external factors should also be considered to make the use of rubrics more effective in performance assessments.

This study illustrated an alternative methodology to estimate variance components and the effects of fixed factors within the same analysis. As with generalizability theory models, crossed random-effects models can compute variance components from crossed facets. At the same time, crossed random-effects models are able to estimate fixed effects related to the facets or other components. A big advantage of this modeling approach over generalizability theory is that it allows for the separation of random and fixed effects from each other. In this study, the effects of rubric use and teaching experience were estimated as fixed effects after all sources

of variation were taken into account. This enabled the researchers to draw additional conclusions about the direct effects of rubric use and teaching experience.

Although the findings of this study enrich the limited knowledge about the effects of rubric use and teaching experience on teachers' scoring behaviors, further research is needed to understand the reasons why these factors are influential. Future research that examines what factors influence teachers' opinions about assigning scores to performance tasks can provide more information about differences in the scoring behaviors of teachers. Researchers should also explore the effects of different types of rubrics, because this study only focused on the use of analytic rubrics for scoring performance tasks. The interaction between teachers and scoring rubrics may also differ based on the type of scoring rubrics.

References

- Ang-Aw, H.T. & Goh, C.C. (2011). Understanding discrepancies in rater judgment on national-level oral examination tasks. *RELC Journal*, 42(1), 31-51.
- Baird, J., & Mac, Q. (1999). *How should examiner adjustments be calculated? - A discussion paper*. AEB Research Report, RC13.
- Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, 5, 27-30.
- Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>.
- Brennan, R. L. (2003). *CASMA research report: Coefficients and indices in generalizability theory* (Research Report No. 1). Iowa City: The University of Iowa Center for Advanced Studies in Measurement and Assessment.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, 24(4), 339-353.
- Brualdi, A. (1998). Implementing Performance Assessment in the classroom. *Practical Assessment, Research & Evaluation*, 6(2). Retrieved August 29, 2012, from <http://PAREonline.net/getvn.asp?v=6&n=2>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Crocker, L., and Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich College Publishers: Philadelphia.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16, 523-545.

- Dochy, F., Gijbels, D., & Segers, M. (2006). Learning and the emerging new assessment culture. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present, and future trends* (pp. 191-206). Oxford, Amsterdam: Elsevier.
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25, 155-185.
- Eliasziw, M., Young, S. L., Woodbury, M.G., & Fryday-Field, K. (1994). Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. *Physical Therapy*, 74, 777-788.
- Farr, R., & Tone, B. (1998). *Le portfolio, au service de l'apprentissage et de l'évaluation*. Montréal/Toronto: Chenelière/McGraw-Hill.
- Fox, J. (2002). *An R and S-PLUS companion to applied regression*. Thousand Oaks, CA: Sage.
- Friel, S. N., Curcio, F., & Bright, G. W., (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education*, 32(2), 124-158.
- Hadden, B. L. (1991). Teacher and nonteacher perceptions of second-language communication. *Language Learning* 41(1): 1-20.
- Hafner, J., & Hafner, P., (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528.
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241-278). Norwood, NJ: Ablex.
- Jolliffe, F. R. (1991). Assessment of the understanding of statistical concepts. In D. Vere-Jones (Ed.), *Proceedings of the third international conference on teaching statistics* (Vol. 1, pp. 461-466). Voorburg, The Netherlands: International Statistical Institute.
- Kulm, G., & Malcolm, S. (1991). *Science assessment in the service of reform*. Washington, D. C.: American Association for the Advancement of Science.
- Lumley, T (1998). Perceptions of language-trained raters and occupational experts in a test of occupational English language proficiency. *English for Specific Purposes*, 17(4), 347-67.
- Messick, S. (1996). Validity of performance assessments. In G. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1-18). Washington, DC: National Center for Education Statistics.
- Meyer, L. (2000). *Lingering doubt examiners: Results of pilot modeling analyses, summer 2000*: AEB Research Report.

- Myford, C.M., & Mislevy, R. J. (1994) *Monitoring and improving a portfolio assessment system*. Princeton, NJ: Educational Testing Service.
- O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49, 14-19.
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research & Evaluation*, 13(4), 1-11. Retrieved on September 29, 2012, from <http://pareonline.net/pdf/v13n4.pdf>
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Pinot de Moira, A. (2003). *Examiner background and the effect on marking reliability*. AQA Research Report, RC218.
- R Development Core Team (2012). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, <http://www.R-project.org>.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Royal-Dawson, L., & Baird, J. (2009). Is teaching experience necessary for reliable scoring of extended English questions? *Educational Measurement: Issues and Practice*, 28(2), 2-8.
- Routman, R. (1991). *Invitations*. Portsmouth, NH: Heinemann.
- Schafer, W., Swanson, G., Bene, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, 14(2), 151-170.
- Shavelson, R. J., & Webb, N. M. (2005). Generalizability theory. In Green, J. L., Camilli, G. & Elmore, P. B. (Eds.), *Complementary Methods for Research in Education*. (3rd ed.) Washington, DC: AERA.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- Shohamy, E., Gordon, C., & Kramer, R. (1992). The effect of raters' background and training on the reliability of direct writing tests. *Modern Language Journal*, 76, 1, 27-33.
- Slater, T. F., & Ryan, J. M. (1993). Laboratory performance assessment. *The Physics Teacher*, 31(5): 306-309.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved on September 11, 2012, from <http://PAREonline.net/getvn.asp?v=9&n=4>

- Stuhlmann, J., Daniel, C., Dellinger, A., Denny, R. K., & Powers, T. (1999). A generalizability study of the effects of training on teachers' abilities to rate children's writing using a rubric. *Journal of Reading Psychology, 20*, 107-127.
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14-23.
- Ward, D. G. (1986). Factor indeterminacy in generalizability theory. *Applied Psychological Measurement, 10*, 159-165.
- Weigle, S. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative & qualitative approaches. *Assessing Writing, 6*(2), 145-178.
- West, B. T., Welch, K. B., & Gallechki, A. T. (2007). *Linear mixed models. A practical guide using statistical software*. Boca Raton: Chapman & Hall/CRC.
- Wiener, R. B., & Cohen, J. H. (1997). *Literacy portfolios: Using assessment to guide instruction*. New Jersey: Prentice Hall.
- Wigglesworth, G. (1994). Patterns of rater behaviour in the assessment of an oral interaction test. *Australian Review of Applied Linguistics, 17*(2), 77-103.
- Wood, R. (1968). Objectives in the teaching of mathematics. *Educational Research, 10*, 83-98.

Çapraz Random Etki Modelleme: Rubrik kullanımı ve Öğretmen deneyiminin Performans Değerleme üzerindeki Etkisinin İncelenmesi

Özet

Atf:

- Kan, A., & Bulut, O. (2014). Crossed random-effect modeling: examining the effects of teacher experience and rubric use in performance assessments. *Eurasian Journal of Educational Research, 57*, 1-28. doi: dx.doi.org/10.14689/ejer.2014.57.4

Problem Durumu

Performans değerlendirme öğrencinin belli bir konuda ne bildiğini ve ne yapabileceğini belirlemek için alternative bir metot olarak görülmektedir. Performans değerlemenin en önemli yetersizliklerinden biri puanlayıcıların puanlamaları arasındaki tutarsızlıklar ve subjektifliktir. Performans değerlemelerin en çok kritik edildiği noktaların başında subjektif puanlama prosedürleri gelmektedir. Performans değerlemelerin etkililiği büyük oranda öğretmenlerin ve rubric puanlama yönergelerinin kalitesi ve koordinasyonuna (etkileşimine) bağlıdır. Öğretmenler ve

rubric puanlama yönergeleri arasındaki etkileşimi daha iyi anlayabilmek için puanlama prosedürlerini etkileyebilecek potansiyel faktörler dikkatlice belirlenmeli ve incelenmelidir. Bu faktörlerden birisi belki de en önemlisi öğretmenlerin mesleki kıdem ve tecrübeleridir. Bu aşamada deneyimli öğretmenlerin öğretim ve değerlemedeki deneyimlerinden dolayı öğrencilerin performansını daha objektif değerlemeleri beklenebilir. Tabii ki bu az tecrübeye sahip öğretmenlerin tutarsız puanlama yaptıkları anlamına gelmez. Öğretmen (özellikleri) performans değerlemenin etkileşimini daha iyi anlayabilmek için tecrübe vb gibi öğretmenle ilgili faktörleri ve performans değerlendirme yaparken öğretmen-rubrik puanlama yönergesi etkileşiminin nasıl olduğunu incelemek hayati öneme sahiptir.

Araştırmanın Amacı

Bu araştırmanın biri pratik diğeri metodolojik olmak üzere iki amacı vardır. Bu çalışma ile (1) performans değerlemede rubrik puanlama yönergesi kullanımı ve öğretmen deneyiminin puanlama üzerindeki etkisini belirlemek ve (2) varyans bileşenlerini kestirmek ve sabit (fixed) faktörlerin etkisini aynı analiz içinde görebilmek için alternatif bir metodu (crossed random effect ve mixed modeller) örnekleyerek tanıtmak amaçlanmıştır.

Metot

Araştırma Deseni

Bu çalışmanın amacı farklı durum ve koşullarda öğretmen deneyiminin ve rubric kullanımının öğretmenlerin öğrenci performans değerlemeleri üzerindeki etkisini incelemektir. Bu amaç doğrultusunda 8. Sınıf öğrencilerine bir performans görevi verilmiş ve onların performansları 17 öğretmen tarafından rubric kullanarak ve kullanmadan farklı zamanlarda puanlanmıştır. Toplanan veriler üzerinde öğrenci görev ve öğretmenin ve faktörlerin etkileşimlerinin öğrenci puanlarındaki değişime katkısının ne olduğunu belirlemek üzere Çapraz random etki modelleme yaklaşımı kullanılmıştır. Buna ek olarak rubric kullanımı, deneyim ve değerlemeler arasındaki zaman model içerisinde sabit etki olarak tanımlanmıştır.

Çalışma Grubu ve Verilerin Toplanması

Bu çalışma 50 adet 8. Sınıf öğrencisi ve sekizinci sınıfta derse giren 17 matematik öğretmeni üzerinde yürütülmüştür. Öğretmenlere ait örneklem 1 ile 26 yıl arasında değişen tecrübeye sahip öğretmenlerden oluşmaktadır. Öğretmen, öğrenci ve performans görevi çapraz yüzeyler olarak tanımlanmıştır. Diğer bir deyişle örnekleme giren tüm öğrenciler aynı sorulara cevap vermişler ve tüm öğrencilerden elde edilen cevaplar 17 öğretmenin herbiri tarafından değerlendirilmiştir. Performans değerlemenin kapsamını grafik yorumlama becerisi oluşturmaktadır. Grafik yorumlama becerisi literature göre çevirme, yorumlama ve sonuç çıkarma (Friel, Bright, & Curcio, 2001; Wainer, 1992) olmak üzere üç alt boyutta tanımlanarak bu boyutlara göre öğrencilerin grafik yorumlama becerisini değerlemek üzere matematik öğretmenleri ve ölçme ve değerlendirme uzmanı araştırmacılar tarafından performans görevleri yapılandırılmıştır. Performans görevi sayısal verilerin grafiksel gösterimini anlama ve ondan yorum çıkarmaya dönük olarak

tasarlanmış dört adet açık uçlu sorudan oluşturulmuştur. İlk iki soru bar grafiği diğer iki soru ise çizgi grafiği üzerine kurgulanmıştır.

Öğrenciler performans görevini tamamladıktan sonra 17 öğretmenin her birinden öğrencilerin dört soruya verdikleri cevapları birbirinden bağımsız bir biçimde değerlemeleri istenmiştir. Soruların puan değerleri birbirine eşit ve herbir soru 25 puan değerindedir. Öğretmenler öğrenci cevaplarını 4 farklı zamanda ve durumda değerlemiştir. İlk durumda öğretmenlerden öğrenci cevaplarını herhangi bir puanlama anahtarı kullanmadan kendi kriterlerine göre puanlamaları istenmiştir. İlk puanlamadan üç hafta sonra öğrencilerin cevapları aynı öğretmenler tarafından bu sefer ekte sunulan rubrik puanlama yönergesi kullanılarak puanlanmıştır. Zamanın puanlamalar üzerindeki etkisini görmek üzere bu işlem ilk iki okumadan 10 hafta sonra tekrar edilmiştir. Öğretmenlerin bir önceki puanlamada herbir öğrenciye verdiği puanı hatırlamasını önlemek için öğrenci cevapları her puanlama durumu öncesinde karıştırılmış ve her öğrenciye her puanlamada farklı bir kod numarası verilmiştir. Böylece öğretmenlerin tekrarlı puanlamaları üzerinde oluşabilecek olası önceki puanlamaların etkisini ortadan kaldırmak amaçlanmıştır.

Verilerin Analizi

Bu çalışmanın deseni birey, görev ve puanlayıcı olmak üzere çapraz üç random yüzeyden oluşmaktadır. Buna ek olarak görev ve puanlayıcılarla çaprazlanmış durum yüzeyi de değişkenliğin diğer bir kaynağı olarak ele alınmıştır. Bu değişkenlik kaynaklarının öğrenci puanları üzerindeki etkisi çapraz random etki (crossed random effect) modeli ile incelenmiştir. Tüm yüzeyler ve etkileşimleri model içerisinde random olarak ele alınmıştır. Random etkilerin yanında öğretmenlerin rubric kullanımı ve deneyimleri gibi iki kestirici de sabit etki olarak ele alınmıştır.

Araştırmanın Bulguları

Araştırma sonuçları, rubric puanlama yönergesinin kullanılmadığı durumlarda öğretmenlerin kendi puanlama kriterlerini oluşturarak performans görevlerini tutarsız bir şekilde puanladıklarını göstermiştir. Fakat öğretmenler iyi tanımlanmış bir rubrik puanlama yönergesi kullandıklarında puanlayıcılar arası güvenilirliğin oldukça yükseldiği gözlenmiştir. Ayrıca tecrübeli öğretmenler ve tecrübesi az olan öğretmenlerin farklı puanlama yaptıkları ve puanlamalarındaki katılık düzeylerinin de farklı olduğu belirlenmiştir.

Araştırmanın Sonuç ve Önerileri

Bu araştırmanın sonuçlarına göre; (1) öğretmenlik tecrübesine sahip öğretmenlerin tecrübesi az olan öğretmenlere göre performans görevlerini daha hoşgörülü puanladıkları, (2) Öğretmenlerin deneyimlerine bağlı olarak oluşan puanlama farklılıklarının öğretmenlerin tamamının rubric puanlama yönergesi kullanması durumunda büyük ölçüde ortadan kalkarak önemini yitirdiği belirlenmiştir.

Bu çalışma ayrıca varyans bileşenlerini kestirmek ve sabit (fixed) faktörlerin etkisini aynı analiz içinde görebilmek için alternative bir metodu da örnekleyerek

tanıtılmaktadır. Bu modelin Genellenebilirlik teorisine göre en büyük avantajı, random ve sabit (fixed) etkilerin birbirinden ayrı ele alınmasına olanak vermesidir.

Öğretmen tecrübelerine ek olarak rubrik kullanımına ve performans değerlendirmeye etki edebilecek diğer dışsal faktörlerinde gözönüne alınması ve incelenmesi önerilebilir. Ayrıca bu çalışma öğretmen tecrübeleri ve rubric kullanımının öğretmenlerin puanlamaları üzerindeki etkisine ilişkin olan sınırlı bilgiyi zenginleştirmesine rağmen özellikle bu faktörlerin puanlama üzerinde neden etkili olduğuna ilişkin yeni araştırmalara ihtiyaç vardır.

Appendix: Scoring Rubric of the Performance Assessment

Performance Categories	Performance Levels				
	0	1	2	3	4
Interpretation of axes	No attempt made.	Does not identify what each axis represents or mistakes the meaning of the axes.	Identifies only one of the axes. Either does not know the other axis or misinterprets it.	Identifies the meaning of each axis but sometimes confuses them.	Is able to identify what each axis represents, and interprets them without any mistake.
Interpretation of intersections	No attempt made.	Interprets the crossing point of x and y elements differently from the correct meaning of the graph.	Is able to identify only a few crossing points of x and y axes, and interprets them correctly.	Is able to identify most of the crossing points on the graph and interprets them correctly.	Is able to identify all crossing points of x and y axes, and interprets them correctly.
Comparisons	No attempt made.	Does not make any comparison based on the values and intervals on the graph or makes incorrect comparisons that do not apply to the graph.	Does not make a detailed comparison of the values and intervals on the graph, and a few comparisons are made correctly.	Is able to make most of the comparisons based on the values correctly, and is aware of the relationship (e.g. increase vs. decrease, small vs. big) between two facts.	Is able to make detailed and meaningful comparisons of the values and intervals based on the graph, and can interpret the relationship between two facts (e.g. increase vs. decrease, small vs. big) correctly.

Integration of facts	No attempt made.	Is not able to relate the facts with the values on the graph. Mentions totally irrelevant facts and events.	Is able to identify only a few data points but cannot explain the big picture by looking at the tendency of the values on the graph.	Is able to identify most data points correctly but cannot fully integrate the facts within the same context. His/her opinions about the facts are not completely clear.	Is able to identify the relationship between the values based on the trend on the graph. Can explain all facts and results consistently within the same context.
----------------------	------------------	---	--	---	--

Graduate Students Perceptions' on Multicultural Education: A Qualitative Case Study¹

Hasan AYDIN*

Betül TONBULOĞLU**

Suggested Citation:

Aydin, H. & Tonbuloğlu, B. (2014). Graduate students perceptions' on multicultural education: a qualitative case study. *Eurasian Journal of Educational Research*, 57, 29-50, <http://dx.doi.org/10.14689/ejer.2014.57.3>

Abstract

Problem statement: The main responsibility in the implementation of multicultural education, which includes notions like equality, respect, and peace, as well as an equal opportunity for success for all students, belongs to teachers. The teachers' perception of and attitude towards multicultural education are directly related to how they will implement multicultural education. Some past studies indicate that there is an undeniable effect of culture on education, yet teaching candidates and managers often don't have sufficient knowledge of multicultural education. Data gathered from this study is expected to contribute to the literature on multicultural education practices expected to be included in Turkey's educational programs.

Purpose of Study: The purpose of this study is to find out the perception of and the evaluation of multicultural education by participants who work as primary and secondary education teachers who are continuing their doctorate level education in curriculum and instruction.

Method: In this inquiry, a qualitative case study approach was used. It has been carried out in the setting of a doctorate level multicultural education class at one of largest state university in Turkey's Educational Science Department. The participants in this study consisted of nine doctorate students taking this course in the Spring of the 2013 academic year, five of which were teachers and four were academicians. Data was collected based on informal interviews, observations, field-notes, and document analysis. A weekly discussion board of reflections, examination papers, and observation forms were also used for more objective results. Gathered data was

¹ This study was presented at the 3rd Congress of National Education on Curriculum and Instruction held on 7-9 May 2014 in Gaziantep University

*Correspondin author: Dr. Yildiz Technical University, Faculty of Education, Curriculum and Instruction Dept., e-mail: aydinh@yildiz.edu.tr

** PhD Student, Yildiz Technical University, Faculty of Education, Curriculum and Instruction Dept, Istanbul, e-mail: betultonbuloglu

evaluated through the “content analysis” technique and the results are presented as a report.

Findings and Results: The findings of this study indicate that equality, democracy, and justice are the most emphasized notions and it has been stated that multicultural education will reinforce these notions. In addition to this, all participants stressed the necessity of taking individual differences into account and expressed that this is cultural richness. Tolerance, respect, and peace are also some of the most commonly mentioned notions.

Conclusions and Recommendations: The need for people with different characteristics and cultures to live with equal rights and every group of the society to have an equal and democratic structure make multicultural education a necessity. Even though the number of studies on multicultural education conducted in Turkey is limited, we recommend that more studies should be conducted like they are in countries with more advanced democracies and that the educational programs should be extended to include multicultural education.

Keywords: multicultural education, equality, democracy, justice, and qualitative research

Introduction

With the disappearance of boundaries as a result of globalization and the immigration to developed countries in the 21st century, the unitary nature of nation-states has been shaken to its core this leading to the transformation of many of them into multicultural states. This societal transformation has also been observed in the field of education, and the number studies on multicultural education, most widely defined as an idea, an educational reform, and as a process (Banks, 2010), has increased considerably, especially in the last few years (Cirik, 2008).

In addition, the goal of multicultural education is to prepare the students for the world with an ever-increasing diversity (Gaff, 1992; Morey & Kitano, 1997), and to increase the students’ capacity for communicating with others by making them participate actively in the process of learning and by endorsing their experiences in a safe classroom setting (Gay, 2000). Effective teaching in a multicultural classroom necessitates cultural sensitivity strategies and the creation of equal opportunities for academic success and personal development for all students (Sharma, 2011). According to Parekh (2002), multicultural education involves intellectual curiosity, self-criticism, the ability to form an independent opinion by evaluating arguments and proofs, respect toward others, sensitivity for other ways of thinking and living, and activities aimed at moving away from an ethnically oriented mentality. Moreover, it is very important to establish a respect towards different cultures and heightened awareness of individuals by creating a suitable setting for mutual understanding in order to promote multiculturalism (Kaya & Aydin, 2014).

The main responsibility in the implementation of multicultural education, which includes notions like equality, respect, and peace, as well as an equal opportunity for success for all students, belongs to teachers. The teachers’ perception of and attitude

towards multicultural education are directly related to how they will implement multicultural education, in other words, how they will integrate the differences in the classroom into the educational process around common values (Sharma, 2011). Teachers' roles are also important in order for students to have high expectations in addition to the proper implementation of multicultural education. Studies have shown that teachers' expectations of their students affect the success of the students. Based on this, it is possible to conclude that the teachers' knowledge of and positive attitude towards multicultural education will allow them to reach a target of high success levels for all students.

To become an effective multicultural educator and teacher, a teacher must examine and continuously transform himself/herself. Schlosser (1992) states that the most effective teachers those who can learn the culture of their students and those that students trust. Gorski (2000) insists that teachers have the responsibility to reassess their own prejudices, partiality, and perceptions that can affect the learning experiences of their students.

In this context, the purpose of this study is to examine the perceptions and evaluation of multicultural education of graduate students who work as primary and secondary education teachers as well as people who are continuing their doctorate level education in curriculum and instruction. Some past studies indicate that there is an undeniable effect of culture on education, yet teaching candidates and managers often do not have sufficient knowledge of multicultural education. Data gathered from this study is expected to contribute to the literature on multicultural education practices expected to be included in Turkey's educational programs.

The research question is, "What are the perceptions and evaluation of multicultural education of graduate students who work as primary and secondary education teachers in Turkey?"

Notions of Multiculturalism and Multicultural Education

The notion of multiculturalism includes the awareness of race, ethnic background, language, sexual orientation, gender, age, disability, social class, education, religious orientation and other cultural dimensions (APA, 2002). Aydin (2013) states that, even though there are multiple forms of multiculturalism, it is, in general, an alternative to the "assimilation" mentality in that it recognizes the existence of different cultures in a country, and that this situation signifies a society that creates room for different cultures to grow on their own. To recognize and value the diversity, rather than ignore it, and to not put all the groups in the same level is a better approach than the contrary (Reitz, 2009).

It is possible to find many different definitions of multicultural education by different theorists working in the field (Gay, 1994). With its most widespread definition, multicultural education is considered as an idea, an educational reform, and as a process (Banks, 2010). Basbay and Kagnici (2011) define multicultural education as a process of building the learning and teaching in a way that promotes cultural pluralism, whereas Amney-Dixon (2004) defines it as an educational approach based on democratic values that promotes cultural pluralism. Mitchell

(1999) gives the definition of multicultural education as educational efforts aimed at creating the necessary positive values for social pluralism by increasing the learning potential of all students, whereas Banks (2010) prefers to define it as education aimed at recognizing and instituting respect and tolerance for differences of age, sexual orientation, disability, social class, ethnicity, color, religion, language and cultural characteristics. Gay (1994) defines multicultural education as an educational policy that has specific values and rules, based on educational rights to ethnic and cultural differences, and that gives equal academic opportunities of success to students. Based on these definitions, it is possible to define multicultural education as a process of creating equal educational opportunities that ensure cultural pluralism within the frame of democratic values by rejecting all kinds of racism and assimilation and respecting all differences. What is meant by multicultural education is not the specialization of formal education for each cultural group, but the sensitivity to multiculturalism in basic education and attention given to diversity by considering multiculturalism in the educational methods (Basbay & Kagnici, 2011).

Strengths and Weaknesses of Multicultural Education

Ameny-Dixon (2004) and Gay (2003) highlights the strengths of multicultural education as follows:

- Emphasizes notions like democracy, equality and justice.
- Prevents intergroup conflicts by ensuring the coexistence of different groups of the society and creates a societal structure in which people live in peace.
- Increases productivity and ensures the intellectual and moral development of all humans because it brings different mental resources together.
- Increases the creative problem-solving skills by applying many different perspectives for the solution of a problem. Increases the positive relations between people by the achievement of common goals and values. Decreases prejudice by bringing different individuals together and letting them interact. Revitalizes societies with the richness of different cultures and helps create a refined world-view.

Researchers who oppose multicultural education have, even though none have a rigid attitude, pointed out some problems arising from multicultural education (Codding & Bergen, 2004). Several researchers argue the weaknesses of multicultural education as follows:

- The most criticized point of multicultural education in the literature is the lack of an agreed upon definition. The most definition by the proponents of multicultural education is Banks's (2008) definition stating that it is a notion, a reform movement, and a process. Since this definition is overly general, it creates a weakness allowing the practitioners to interpret it as they see fit. Some negative results arising from the practice can result instead of the positive objectives (Thomas et al., 1994).
- Many researchers, while studying multicultural education, emphasized race and ethnicity and neglected other differences like gender, religion, language and social class (Furman, 2008).
- Many studies conducted on teachers who are the implementers of multicultural education have shown that they do not believe in multicultural education (Gay,

2003), that they cannot find enough support from the institution they work in (Van Hook, 2002), and that they do not have sufficient knowledge on how to practice multicultural education.

- Overemphasizing multiculturalism creates an excessive ethnical awareness (Webster, 1997). Even though individuals have more than a single identity, the overemphasis on ethnic and racial identities makes these identities dominant and causes the individual to distance themselves from the mainstream culture and this constitutes a risk against the national integrity and social order (Glazer, 1997).
- Multiculturalism is a notion based on cultural relativism and claims that no culture is superior to another. But every culture considers its own values, norms, points of view and beliefs as superior than the others' (D'Souza, 1995).

Despite these dissenting opinions surrounding it, multicultural education, especially in the US, has come a long way in terms of curriculum (Banks, 2008). Establishment of a multicultural mentality in the society is very important for maintaining the social order and protecting the national integrity and unity.

Teachers' Role in Multicultural Education

Basbay and Kagnici (2011) state that, alongside those who consider multiculturalism as a positive and passive process, there are those who criticize multiculturalism by claiming that it can cause the deterioration of the social structure. Just like with all practices related to education, teachers have an important role in understanding and implementing multicultural education. Gay (1994) states that teachers generally perceive their own values, beliefs, and experiences as standard rules and that what they know and what they teach is governed by educational principles. She states that they may engage, knowingly or unknowingly, in erroneous educational practices because they do not understand the racial, social, and linguistic diversity of their students. Formal education, as Costa (1997) states, depends on the attitude and professional preparation of the teacher and this is even more important in multicultural classrooms. Pea (1997), in his analysis on teachers' perceptions, states that teachers put too little time into getting to know their students and they know very little about them for this reason. He also states that the experiences of students belonging to cultural minority groups are ignored by the teacher due to the teacher's perception and to the fact that he/she may not have taught similar students before. Aydin (2012) emphasizes the importance of training educators who can design educational programs starting from preschool in the establishment of an academic infrastructure that will implement a multicultural education system.

Washington (2003) states that teachers need three important basic skills, namely "understanding oneself", "understanding the culture of others," and "academic-multicultural qualifications". According to Basbay and Kagnici (2011), of the qualifications determined by the Ministry of National Education, those related to knowing the student, diversifying the education by taking individual differences into consideration, valuing, understanding and respecting the students and attaching importance to national and universal values necessitate that the teachers be at a certain level in terms of multiculturalism. NCATE 2006 attaches importance to

teacher candidates' abilities to develop classroom settings by taking diversity into consideration and unifying the diversity; this document contains articles necessitating the teachers to possess these qualifications.

Polat (2009) states that the objective of education is to focus on, instead of to standardize, the student, accepting him/her as he/she is and developing his/her abilities and that, consequently, students need educators who consider multiculturalism natural and who can provide a multicultural education. Cirik (2008) emphasized that the main principles of learning and teaching should defend students against to racism, prejudice, sexism and other forms of discriminations Gorski (2000) indicates a strongly that teachers have a responsibility to continuously examine their own prejudices, partiality, and perceptions that can affect students' experiences. To become an effective multicultural educator and teacher, the teacher has to keep examining and transforming himself/herself. According to Bigatti et al. (2012), being a multicultural teacher involves more than presenting course content on diversity. It necessitates ensuring an inclusive and progressive classroom setting, feeling the communication between the students, and facilitating the students' ability to learn with the help of different educational techniques and evaluations (Gay, 2004b). Aydın (2013) states that educators can develop the students' ethnic, linguistic and cultural literacy, support their self-esteem and development, teach them the importance of inclusiveness and tolerance, encourage them to work and interact with people who are different than them and provide a more effective education by these means. Based on these statements one can conclude that teachers have a critical role in the conception and implementation of multicultural education based settings and programs.

Method

Research Design

In this study, the researchers used a qualitative case study approach. According to Bogdan and Biklen (1992), a case study is the detailed analysis of an environment, a single document, or a special event. Denzin and Lincoln (2011) have stated that, despite the vastness of statistical methods, the main difficulty of the case study is to come up with a study that is detailed, rich, and complete, and that includes all case variances. The study has been conducted in the scope of the multicultural education class provided by the curriculum and instruction department at one of the largest state universities in Turkey. The study group was selected with purposeful sampling. According to Erdogan and Yazicioglu (2007), in purposeful sampling, the sample is selected from the population by determining the units that represent the needed data. Necessary human subject permissions were obtained from the Social Sciences Institute (SBE) of the university.

Research Sample

The participants of this study consist of nine doctoral students taking the multicultural education class in spring 2012-2013 academic years, five of which are teachers and four are academicians (T1, T2, T3, T4, T5 & A1, A2, A3, A4). All of the participants selected the Multicultural Education class voluntarily and gave written permission for the utilization of information gathered from them as academic data.

The data were collected in the scope of the volunteering principle. Participants consist of five female and four male, and their average age is 31.

Research Instrument and Procedure

In this study, the data was collected through “triangulation” of multiple data collection methods. Triangulation overcomes the limitations of one data collection method by using another data collection method in conjunction and thus increases the validity and reliability of the findings (Yildirim & Simsek, 2008).

Interview. Over the course of the study, the students were asked structured and unstructured questions. The questions were mostly aimed at finding out their points of view, attitudes, and opinions concerning multiculturalism and multicultural education. All communications and questions’ answers have been written by the researchers during rest of semester which covers fifteen weeks.

Weekly Discussion Board. For seven out of a total of fifteen weeks that the Multicultural Education class was taught, articles published on a national and international scale about multicultural education were sent every week to students by email. Every student was asked to read these articles and write an essay in accordance with the questions sent. These essays were posted on the discussion board and the multicultural education class was structured around these discussions. A total of 26 articles, six articles for the first week, three for the second week, two for the third week, seven for the fourth week, four for the fifth week and four for the sixth week were sent, all relating to the subjects to be handled. The students were asked to analyze these and write their essays accordingly.

Examination papers and students course materials. An examination consisting of 14 structured interview questions was conducted to measure and evaluate the students’ multicultural education perceptions. To ensure the validity and reliability of the examination questions, three experts were consulted and the examination was conducted after the questions were structured. The students were informed before the examination that the first 12 questions were interpretation questions and that they did not have any evaluation purpose and that the last two questions were questions of knowledge and that they would be evaluated accordingly.

Observation Form. Students’ attitudes were observed throughout the period and the changes in their statements and behaviors were recorded on the observation form. Particularly, their initial knowledge, attitude and prejudices and the positive and negative changes during the lectures have been recorded.

Data Analysis

All interviews were taped and transcribed after importing to the digital computer media. Qualitative data were analyzed with the *content analysis* method and the results were made into reports. Content denotes what is contained and content analysis is the analysis of what is contained in a message. Broadly content analysis may be seen as a method where the content of the message forms the basis for drawing inferences and conclusions about the content (Lal Das, & Bhaskaran, 2008; Nachmias & Nachmias, 1976)..Further, content analysis falls in the interface of observation and document analysis. It is defined as a method of observation in the sense that instead of asking people to respond to questions, it “takes the

communications that people have produced and asks questions of communications” (Kerlinger, 1973, as cited in Prasad, nd., p.2). Therefore, it is also considered as an unobtrusive or non-reactive method of social research. In addition, discovered themes have been verified with the Nvivo 10 software and their correctness has been confirmed by five peers who enrolled in doctoral program in education.

Observation is a method used to define a behavior emerging in an environment in a detailed manner. The most important property of observation is that it provides first hand data to the researcher (Yildirim & Simsek, 2008). The preferred method of this study, unstructured field study, is conducted in the natural environment of the behavior and it is generally carried out with a method called ‘participatory observation’ in which the researcher participates in the environment. In participatory observation, the observer is with the observees and acts like one of them (Karasar, 2007). In this study, this type of observation was preferred and an observation form was used to facilitate the recording of the data. Document analysis, on the other hand, consists of the analysis of written material containing information on the researched notions and its validity increases when used with other data collection methods like observation and interview through data triangulation (Yildirim & Simsek, 20088). Students’ essays and examination papers were analyzed with the thematic and content analysis method. By taking all the data into consideration, the most emphasized notions by the participants were coded. The most important themes discovered are analyzed and presented.

Role of the Researchers

In this study, the researchers attempt to present the findings of the process as is and strive to not project their own views. Observations were carried out by two researchers and the interviews by one. The documents were analyzed separately by the two researchers, the findings were compared, and a common conclusion was reached with the results then verified with the Nvivo 10 software. After the as-is presentation of the findings, personal opinions and the findings have been handled in the discussions section.

Validity and Reliability

- a) To increase the internal validity (plausibility) of the study, triangulation was used. The findings have been presented based observation, interview, and document analysis. After the taped questions and answers, the students have confirmed their own responses. In addition to this, the students confirmed their views and cleared up misunderstandings by reading their weekly essays in the classroom. The study had a nine week duration, which enabled a long-time interaction with the participants and collection of deep data.
- b) To increase the reliability (consistency) of the study, direct quotations have been used in the presentation of the findings. Two researchers conducted the observations, and these observation findings were compared at the reporting phase in order to reach common conclusions. The researchers read the interview and observation findings and possible codes and themes were thus determined and then verified by the Nvivo 10 software.

Results

The analysis of the examination papers and essays of the participants leads to the observation that all participants agree on the necessity of multicultural education and have a positive point of view toward it. The most emphasized notions are equality, justice, and democracy and all participants agreed that personal differences have to be taken into consideration and that it's richness of cultural and diversity. Tolerance, respect, and peace are also among the most reiterated notions.

Theme One: Equality

The analysis of the documents obtained during the application phase of the research reveals that the word *equality* has been repeated 37 times by the participants and that this is the most used description. Almost all of the nine participants emphasized the notion of equality and stated that multicultural education ensures equality. Five participants defined multicultural education with an emphasis on equality. One participant (A1) defined multicultural education as follows:

Multicultural education is the constitution of the umbrella of the teaching programs in a way to include all kinds of difference and means education *in equal conditions* for every individual regardless of their religion, language, race, gender and physical and mental differences.

Another participant's (A3) definition of multicultural education emphasizes again the notion of equality:

Multicultural education is, in the most general sense, offering *equal chances of success* to students having different races, religions, languages, cultures etc.

All nine participants stated that the biggest advantage and strength of multicultural education is 'ensuring the equality' and that this also brings peace. For example one participant (A1) who considers equality as the greatest gain of multicultural education explained:

The biggest advantage of multicultural education is the mention of all individuals of the society in the educational system and the extension of *an egalitarian approach* to all schools.

Similarly, another participant (T4) expressed the opinion that the biggest advantage of multicultural education is to ensure the equality with the following sentence:

I think its biggest advantage is not to abandon students coming from disadvantaged groups of the society to their fate and support them instead and thus ensure them *an equal right of education*.

Some participants stated that if they prepared a multicultural program, they would promote the ability for all groups to express their opinion equally and emphasized its importance. One participant (T2) said

The program would contain points like common education in two languages, reduction of prejudices and *equal right* to express their opinion for all peoples" in this regard while another stated "I would prepare the

curriculum in such a way that would represent different groups of the society *equally*.

Yet another participant (A2) expressed his opinion about the content of the program as follows:

Respect to differences, equality of opportunity in education and the contents of elements like democracy and social justice should be handled abundantly educational programs.

The participants also mentioned the factor of equality in the list of descriptions that can be done to promote and popularize multicultural education. One participant stated the need for “the school personnel to receive training emphasizing the advantages of multiculturalism in school including notions like diversity, culture, *equality*, human values etc.” and another emphasized “a systematic, *equal* and organized multicultural education has to be legislated first”.

All participants stated that different strategies and methods can be used for different social classes and stressed its necessity. They pointed out that the classes taught should also contain the notion of equality and gave examples. One participant (T5) expressed his/her opinion as follows:

For different classes, many elements can be included in the learning environment. Traditions and special days of different religions (Ramadan Feast, Christmas Day etc.), different ethnic cultures (blacks, whites, Asians, Kurds, Turks etc.), all kinds of activities against prejudice and emphasizing gender equality should take their place in the classroom.

Another participant (T1) listed what can be done for different classes as follows:

In Project based learning environments, each student can prepare projects and homework projecting their own mother tongue and culture. The UK is performing this very well, so can we. Also, we can stop using civics classes as a tool for assimilation and transform it into a higher quality class where universal values like democracy, pluralism, social justice and equality.

A general analysis of the answers provided by the participants reveals that the notion of equality is stressed and that there is a consensus about multicultural education’s importance in ensuring equality.

Theme Two: Democracy

The analysis of the documents obtained during the application phase of the research reveals that the word *democracy* was used 29 times by the participants and it is one of the most often repeated notions. All nine participants emphasized the notion of democracy and talked about the importance of multicultural education in forming a democratic society. One participant cited the immaturity of Turkish democracy as one of the obstacles before the application of multicultural education and two participants emphasized the notion of democracy while defining multicultural education. These definitions were worded as follows:

Multicultural education is a *democratic* reform movement where students having many differences like race, religion, language and gender have an equal right of education. (T3)

It is the ability of individuals from different cultures regardless of their languages, religions, races or genders to continue their education in the same learning environment in the scope of *democracy and social justice*. (A4)

Seven of the nine participants mentioned concerns that multicultural education might divide the society and deteriorate its unity and two emphasized assimilation. The comments on these concerns stated that multicultural education would, on the contrary, contribute in the formation of a democratic environment. One participant (T4) expresses his/her opinion as follows:

Multicultural education does not divide a society or a country. On the contrary, in societies having differences, it helps them survive in a more *democratic* framework.

One of the most cited advantages and strengths of multicultural education is its ability to form a democratic environment. One participant (T2) stated;

To be able to apply notions like equality and *democracy* in a healthier way" as an advantage of multicultural education while another claimed "to help the development of sciences like tolerance, brotherhood, rights, law and *democracy*", yet another said "that it ensures a *democratic* social order.

Also, one participant (T5) stated as follow;

Trying to apply it in societies where democracy is not well developed may cause social unrest", mentioning the importance of the existence of a democratic society for multicultural education to be applied.

The opinion that the integration of multicultural education into the Turkish educational system would bring about the development of a consciousness of democracy was suggested and one participant (A2) listed the advantages of this as follows:

There would be an opportunity to benefit from different cultures, a consciousness of *democracy for all* would develop, an environment of trust and peace would be formed away from fear and concerns, it would ensure cultural interaction, it would cause social transformation and progression.

Another participant (A1) expressed his/her opinions on the matter as follows:

I think that a transition to a much more *democratic* educational system taking individual needs and differences into account would be the biggest gain.

When listing what is necessary for multicultural teacher training programs for the Turkish school system, the idea of democracy was emphasized again and it was stressed that the programs must be based on democratic principles. One participant (A3) said:

I would strive to make it a program based on Islamic tolerance, referencing successful historical examples conforming the universal principles of human right and *democracy*.

Another participant (A4) expressed his/her opinion as follows:

Our academicians who will train the teachers have to communicate that education must be carried out in a framework of love, respect, social justice and democracy and they should practice accordingly.

When the participants' statements are analyzed in general, it is revealed that the notion of democracy was stressed quiet often and that multicultural education's contributions to the creation of a democratic society and democratic environment were often reiterated.

Theme Three: Justice

An analysis of the documents obtained during the application phase of the research reveals that the word *justice* was used 17 times by the participants and that it is another of the most repeated notions. All nine participants emphasized the notion of justice and mentioned the importance of multicultural education in ensuring justice. Two participants mentioned this concept while defining multicultural education and two other participants stated that multicultural education is related to the notion of social justice. One of the multicultural education definitions is as follows:

Multicultural education is to accept everyone regardless of their religion, language, ethnic background and gender with their own values. It is the ability to approach everyone with *justice*, tolerance and impartiality. (T5)

About the notions that multicultural education is related to and its role on the establishment of justice, two participants express the following opinions:

Multicultural education is related to notions like difference, equality, *social justice* etc. and it is stated that multicultural education must be practiced for societies living in peace. (A3)

Multicultural education is a theory that first appeared and developed in the US following the Civil War but it is also a quest for a solution to eliminate the *social injustice* in countries like the US, Canada and the UK where the demographics change with the increase of minorities. In this context, multicultural education is related to notions like democracy, citizenship, immigration, differences, *social justice*, minority rights and integrations. (T4)

The fact that multicultural education stresses the notion of justice and that it has a quality that can establish justice in the society was indicated as strength of multicultural education by four participants. One participant emphasized multicultural education's importance for human rights and social justice, another stated that multicultural education infuses the sentiment of equality and justice to a participatory society, yet another mentioned multicultural education's establishment of social justice and yet another discussed its contribution to the sentiment of justice and self-esteem.

Social justice was described as being one of the necessary elements for the implementation of multicultural teacher training programs. One participant (A2) stated:

Respect to differences, equality of opportunity in education and the contents of elements like democracy and social justice should be handled abundantly educational programs.

Another participant (A4), on the matter of teacher training, stated:

Our academicians who will train the teachers have to communicate that education must be carried out in a framework of love, respect, social justice and democracy and they should practice accordingly.

Participants also emphasized the notion of social justice when talking about the steps that can be taken to improve the teacher training programs. One participant (A2) stated:

First of all, teachers should be trained to embrace differences. The most fundamental quality that a teacher has to have is *justice*. In this context, *justice and social justice* have to be included to the program.

A general analysis of the answers reveals that multicultural education's importance in establishing social justice is often stressed and considered very important. All participants had positive thoughts on multicultural education. Most participants emphasized that multicultural education contributes societies with social justice, equality, and democratic values for each individuals. Thus, both teachers and academics should be trained regarding multicultural education for democratic societies.

Observation Findings

At the first week of the study, we asked the students about their ethnic background and religion and we observed that some of them were reluctant to give this information or comment. After the executive of the class defined multiculturalism and multicultural education, introduced the subject and revealed his/her own ethnic background and religion, students started to reveal more information. Next week, students who read the relevant articles and wrote their essay discussed what they understood from the lectures, their concerns about multicultural education, the points that surprised them, the rationale and objectives of multicultural education, the reasons why it is beneficial to students coming from different cultures and its possible benefits for the Turkish educational system. Students most often expressed their concerns that multicultural education could divide the country. However after the readings they performed during the following weeks, we observed that they decided that multicultural education is actually a unifying education design and that it aims at ensuring a more egalitarian, more democratic, and more just social order. All students agreed that multicultural education is necessary and important and that multicultural education is very important for students coming from different cultures to live their own values without being assimilated and to be more successful.

Discussion and Conclusion

Multicultural education is a field of study that helps students acquire the knowledge and skills necessary for participating to the democratic society and those coming from different cultures to have equal opportunity of education (Halvorsen & Wilson, 2010). Analysis of educators' opinions on multicultural education and interpretation of the findings to help reorganize the training programs will contribute to a more equal, more democratic and a more just educational system. The analysis of the findings of this study reveals that all participants state that multicultural education is important and necessary and they emphasized, above all, equality, democracy and justice on the matter of multicultural education.

The study's findings show that multicultural education, besides many other advantages, ensures equality. D'Souza (1995) states that multiculturalism considers all cultures as equals and none of them are superior to another. Nelson (2001) states that valuing diversity in socially organized belief systems, races and cultures is the key to providing equal educational opportunities to all students. ITARI (1991) identifies the central value of multicultural education programs as education of different groups around values like equality, justice, human rights and democracy. Sleeter and Grant (1987) emphasizes that multicultural education movement is a civil rights movement aiming at making education equal for all social groups. Sharma (2011) related an effective education in multicultural classrooms to providing equal opportunities for academic success and personal development. Pang (2001) stresses that multicultural education can be a starting point for the elimination of inequalities in the society due to its decisiveness in treating the deficiencies and mistakes of the current educational system. In this study, in parallel with the literature, the most mentioned notion by the doctorate students was equality and it was stated that multicultural education will contribute to everyone having equal opportunities and equal right to education.

The notion of democracy was also one of the most reiterated notions by the participants. Similarly, there are many studies mentioning the positive contribution of multicultural education to democracy. Cole et al. (2003), define multicultural education as a learning and teaching approach based on democratic values and beliefs and pointed out the cultural diversity in an interconnected world with cultural pluralism. Nieto (1992) defines multicultural education as an educational system reform aiming at providing education without regard to any kind of discrimination to develop the relations between the individuals in the classroom. Polat (2009) indicates that the quality of multicultural education is proportional to the democratic practices of the country. Gay (2003) states that one of the strengths of multicultural education is its emphasis on notions like democracy, equality, and justice. Banks (2008) states that the goal of a multicultural education program is to democratize the education programs in order to provide everyone with a more impartial and free point of view. Similarly, in this study the participants stated that multicultural education would strengthen the democracy in the society and contribute to a democratic learning environment.

Justice is yet another notion that was often emphasized in this study. Aldridge, Calhoun and Aman (2000) state that multicultural education helps the society be more tolerant and more just, and realize that it is richer than any of its constituents.

Herring and White (1995) also state that teachers should use non-sexist language and implement an egalitarian education. According to Hunter (1974), multicultural education is the structuring of the education in the context of equality, mutual respect, acceptance, understanding and moral principles for realizing the democratic ideals, meeting the needs of different groups of the society and ensuring the social justice. Grant (1977b) defines multicultural education as an educational approach deriving all the necessary content for a good education and to better understand the global society from the existing cultures of the students based on social justice for all humans, and the power of alternative life styles, human rights and diversity. According to Banks (2008), one of the main goals of an educational system promoting a multicultural literacy is to help the student develop a democratic and egalitarian society, know, be sensitive and act. In this study, the participants, similar to the findings in the literature, mentioned the importance of multicultural education in ensuring justice, they emphasized the notion of justice in their definitions of multicultural education and related multicultural education to the notion of social justice.

In a way, the need for people with different characteristics and cultures to live with equal rights and every group of the society to have an equal and democratic structure make multicultural education a necessity. Stradling (2003) states that multicultural education creates a possibility to accept that there are different points of view and that they are equal and to understand how the other side feels. Even though the number of studies on multicultural education conducted in Turkey is limited, we recommend that more studies should be conducted like they are in countries with more advanced democracies and that the educational programs should be extended to include multicultural education.

References

- Aldridge, J., Calhoun, C., & Aman, R. (2000). 15 misconceptions about multicultural education. *Focus on Elementary*, 12(3).
- Ameny-Dixon, G. M. (2004). Why multicultural education is more important in higher education now than ever: A global perspective. *International Journal of Scholarly Academic Intellectual Diversity*, 8(1), 1-9.
- APA. (2002). *Guidelines on multicultural education, training, research, practice, and organizational change for psychologists*. Retrieved on December 04, 2013, from <http://www.apa.org/pi/oema/resources/policy/multicultural-guidelines.aspx>
- Aydin, H. (2012). Multicultural education curriculum development in Turkey. *Mediterranean Journal of Social Sciences*, 3(3), 277-286. Doi:10.5901/mjss.2012.v3n3p277
- Aydin, H. (2013). *Dünyada ve Türkiye’de Çokkültürlü Eğitim Tartışmaları ve Uygulamaları*. [Discussions and Practice of Multicultural Education in Turkey and in the world.] Ankara: Nobel Publishing House.

- Banks, J. A. (2008). Çokkültürlü Eğitime Giriş [Introduction to Multicultural Education]. (H. Aydın, Çev.) İstanbul: Anı Yayıncılık.
- Banks, J.A. (2010). Multicultural Education: Characteristics and Goals. J.A.Banks & C.H.M.Banks (Eds), *Multicultural Education: Issues and Perspectives* (pp.3-30) (7th Ed.). ABD: Wiley.
- Başbay, A. & Kağnıcı D.Y. (2011). Çokkültürlü Yeterlik Algıları Ölçeđi: Bir Ölçek Geliştirme Çalışması [Perceptions of Multicultural Competence Scale: A Scale Development Study]. *Eğitim ve Bilim*, 36 (161), 199-212.
- Bigatti, S. M., Gibau, G. S., Boys, S., Grove, K., Ashburn-Nardo, L., Khaja, K., & Springer, J. T. (2012). Faculty Perceptions of Multicultural Teaching in a Large Urban University. *Journal of the Scholarship of Teaching and Learning*, 12(2), 78-93.
- Bogdan, R. ve Biklen, S. K. (1992). *Qualitative Research for Education*. Boston: Allyn and Bacon.
- Brodkin, K. (1998). *How the Jews Became White Folks and What That Says About Race in America*. New Brunswick, NJ: Rutgers University Press.
- Cırık, İ. (2008). Çokkültürlü eğitim ve yansımaları [Multicultural Education And Its Reflections]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34, 27-40.
- Codding, R. & Bergen, K. (2004). *Multicultural education: An Extensive Literature Review*. Retrieved on November 23, 2013, from http://humboldt.edu/sociology/downloads/senior_projects/2004Codding%20Bergen.pdf adresinden erişildi.
- Cole, D., Bennett, C., & Thompson, J. N. (2003). Multicultural teacher education: Implications for critical mass and all-minority classes at a predominantly White university. *Journal of Classroom Interaction*, 38 (1), 17-28.
- Costa, X. B., (1997), *Intercultural education: Theories, policies and practices*. Brookfield, Vermont: Ashgate.
- Denzin, N. K. & Lincoln, Y. S. (2011). *The SAGE Handbook of Qualitative Research*. Thousand Oaks, CA: SAGE Publication.
- Dickeman, M. (1973). Teaching Cultural Pluralism. In J. A. Banks (Ed.), *Teaching Ethnic Studies: Concept and Strategies* (pp. 5-25). Washington, DC: National Council for the Social Studies.
- D'souza, D. (1995). *The end of racism: Principles for a multiracial society*. New York: Free Press.
- Fordham, S. (1988). Racelessness as a Factor in Black Students School Success: Pragmatic Strategy or Pyrrhic Victory? *Harvard Educational Review*, 58(1), 54-84.

- Furman, J. S. (2008). Tensions in Multicultural Teacher Education Research Demographics and the Need to Demonstrate Effectiveness. *Education and Urban Society*, 41(1), 55-79.
- Gaff, J. G. (1992). Beyond politics: The educational issues inherent in multicultural education. *Change*, 24 (1), 30-35.
- Gay, G. (1994). A synthesis of scholarship in multicultural education. Urban Monograph No. RI88062012), OakBrook, Illinois, North Central Regional Educational Laboratory.
- Gay, G. (2000). *Culturally Responsive Teaching: Theory, Research and Practice*. New York: Teacher College Press.
- Gay, G. (2003). The Importance of Multicultural Education. *Association for Supervision and Curriculum Development*, pp.30- 35. Retrieved on December 3, 2013, from http://www.achievementseminars.com/seminar_series_2008_2009/readings/importance_multicultural_education.pdf
- Gay, G. (2004b). Curriculum theory and multicultural education. In J. A. Banks & C. A. Banks (Eds.), *Handbook of research on multicultural education* (2nd ed.) (pp. 30-49), San Francisco: Jossey-Bass.
- Glazer, N. (1997). *We Are All Multiculturalist Now*. Cambridge, Massachusetts: Harvard University Press.
- Gorski, P. C. (1999). A Brief History of Multicultural Education. Retrieved on November 20, 2013, from http://www.edchange.org/multicultural/papers/edchange_history.html
- Gorski, P. (2000). The challenge of defining a multicultural education. Retrieved on November 22, 2013, from <http://www.edchange.org/multicultural/initial.html>
- Grant, C. A. (1977b). *Multicultural education: Commitments, issues and applications*. Washington, D.C.: Association for Supervision and Curriculum Development.
- Halvorsen, A. L. & Wilson, S. M. (2010). Social Studies Teacher Education. In Peterson, P., Baker, E., & McGaw (Eds), *International Encyclopedia of Education* (3rd ed.) (pp.719-737), Oxford: Elsevier Academic Press.
- Herring, R.D., & White, L.M. (1995). School counselors, teachers, and the culturally compatible classroom: Partnerships in multicultural education. *Journal of Humanistic Education and Development*, 34(2), 52-64.
- Hohensee, B. J, & Derman-Sparks. L. (1992). Implementing an Anti-Bias Curriculum in Early Childhood Classrooms. *ERIC Digest*.

- Hunter, W.A. (Ed.).(1974). *Multicultural education through competency based teacher education*. Washington, D.C.: American Association of Colleges for Teacher Education.
- Kaya, İ. & Aydın, H. (2014). *Çoğulculuk, Çokkültürlü ve Çokdilli Eğitim*. [Pluralism, Multicultural, and Multilingual Education]. Ankara: Anı Publishing.
- Kerlinger, F.N. (1986). *Foundations of behavioural research (3rd ed)*. New York: Holt, Rinehart and Winston.
- Kodish, B.I. (2003). What we do with language-what it does with us. *ETC: A review of general semantics*, 60(1), 383-384.
- Lal Das, D.K and Bhaskaran, V (eds.). (2008) *Research methods for Social Work*, New Delhi:Rawat, pp.173-193.
- Lee, K. S.,&Carrasquillo, A. (2006). Korean college students in United States: Perceptions of professors and students.*College Student Journal*, 40(2), 442-456.
- Mitchell, B. M. (1999). *Encyclopedia of multicultural education*. Westport, Conn: Greenwood Press.
- Morey, A. &Kitano, M. (Eds.). (1997). *Multicultural course transformation in higher education: A broader truth*. Boston: Allynand Bacon.
- Nachmias, D. & Nachmias, C. (1976). Content analysis. In *Research methods in the social sciences* (pp.132-139), UK: Edward Arnold.
- Nelson, T. (2001). Editor's Introduction: Advancing educational opportunities in a multicultural society. *Teacher Education Quarterly*, 26(3), 147-156.
- Nieto, S. (1992). *Affirming diversity: The sociopolitical context of Multicultural Education*. New York: Longman.
- Pang, V. O. (2001). *Multicultural education: A caring-centered, reflective approach*. New York: McGraw- Hill Press.
- Parekh, B., (2002), *Çokkültürlülüğü Yeniden Düşünmek, Kültürel Çeşitlilik ve Siyasi Teori*. [Rethinking Multiculturalism- Cultural Diversity and Political Theory]. Çev: Bilge Tanrıseven, Ankara: Phoenix Yayınevi.
- Pena, R. A., (1997). Cultural differences and the construction of meaning: Implications for the leadership and organizational context of schools. *Education Policy Analysis Archives*, 5(10), 1-19.
- Polat, S. (2009). Öğretmen Adaylarının Çok Kültürlü Eğitime Yönelik Kişilik Özellikleri. [Teachers' Personality Traits Towards Multicultural Education] *International Online Journal of Educational Sciences*, 1 (1), 154 – 164.
- Prasad, B., D. (nd). Content Analysis:A method in Social Science Research. Retrieved on December 29, 2013, from <http://www.css.ac.in/download/Content%20Analysis.%20A%20method%20of%20Social%20Science%20Research.pdf>

- Reitz, J. G. (2009). Assessing Multiculturalism as a Behavioural Theory. In Reitz, J. G., Breton, R., Dion, K. K., & Dion, K. L. (Eds.), *Multiculturalism and Social Cohesion: Potential and Challenges of Diversity* (s. 1-47). Toronto, Canada: Springer Science + Business Media Press.
- Schlosser, L. K., (1992). Teacher distance and student disengagement: School lives on the margin. *Journal of Teacher Education*, 43(2), 128-140.
- Sharma, S. (2011). Multicultural education: Teachers' perceptions and preparation. *Journal of College Teaching & Learning*, 2(5), 53-64.
- Sleeter, C. E., & Grant, C. A. (1987). An analysis of multicultural education in the United States. *Harvard Educational Review*, 57(4), 421-445.
- Stradling, R. (2003). 20. Yüzyıl Avrupa Tarihi Nasıl Öğretilmeli?. [How Should be Thought 20th Century European History]. İstanbul: Türkiye Ekonomik ve Toplumsal Tarih Vakfı Yayınları.
- Thomas, D.G., Chinn, P., Perkins, F., Carter, D.G. (1994). Multicultural Education: Reflections on Brown at 40. *The Journal of Negro Education*. 63(3), 460-469.
- Van Hook, C.W. (2002). Preservice Teachers' Perceived Barriers to the Implementation of a Multicultural Curriculum. *Journal of Instructional Psychology*. 29(4).
- Washington, E. D. (2003). The multicultural competence of teachers and challenge of academic achievement. In D. B. Pope-Davis, H. L. K. Coleman, W. M. Liu & R. L. Toporek (Eds.), *Handbook of multicultural competencies in counseling and psychology*. California: Sage Publications.
- Webster, Y.O. (1997). *Against the Multicultural Agenda: A Critical Thinking Alternative*. Westport, Connecticut: Praeger Publishing.
- Woodrow, D., & Sham, S. (2001). Chinese pupils and their learning preferences. *Race Ethnicity and Education*, 4(4), 377-394.
- Yıldırım, A. & Şimşek, H. (2008). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri [Qualitative Research Methods in Social Sciences]*. Ankara: Seçkin Yayıncılık.

Doktora Öğrenimi Gören Öğretmen ve Akademisyenlerin Çokkültürlü Eğitim Algılarının İncelenmesi

Özet

Atf:

Aydın, H. & Tonbulođlu, B. (2014). Graduate students perceptions' on multicultural education: a qualitative case study. *Eurasian Journal of Educational Research*, 57, 29-50, <http://dx.doi.org/10.14689/ejer.2014.57.3>

Problem Durumu: Çokkültürlü eğitim; ırkçılığı ve asimilasyonu reddederek ve her türlü farklılığa saygı göstererek, demokratik değerler çerçevesinde kültürel çoğulculuđu sağlamak için eşit eğitim ve öğrenim fırsatları oluşturma süreci olarak tanımlanabilir. Çokkültürlü eğitimin amaçları genel bir ifadeyle farklı gruplar arasındaki iletişimi geliştirmek, saygı ve hoşgörüyü artırmak, bireylerin kendisiyle barışık olmalarını sağlamak ve daha demokratik, adil bir toplum düzeni oluşturmak olarak sıralanabilir. Çokkültürlü eğitimin temelinde bulunan eşitlik, saygı, barış, vb kavramlarla ve tüm öğrencilere eşit başarı olanađı sağlama amacıyla uyumlu bir şekilde uygulanması için temel görev öğretmenlere düşmektedir. Öğretmenlerin rolü, çokkültürlü eğitimin amaç ve kapsamına uygun bir şekilde uygulamasında olduđu kadar farklı gruplara mensup tüm öğrencilerinden yüksek beklentilerinin olmasında da belirleyicidir. Araştırmalar öğretmenin öğrencilerinden beklentilerinin öğrenci başarısını etkilediđini ortaya koymuştur. Buradan hareketle öğretmenlerin çokkültürlü eğitim ile ilgili bilgi ve olumlu tutum sahibi olmasının, tüm öğrencileri için yüksek başarı hedeflemesini sağlayacağı söylenebilir. Çokkültürlülükle ilgili yapılmış birçok araştırma sonucu, kültürün eğitim üzerinde yadsınamaz bir etkisinin bulunduđunu gösterdiđi halde öğretmen, öğretmen adayı ve yöneticilerin çokkültürlülükle ilgili yeterince bilgi sahibi olmadığını ortaya koyan araştırmalar mevcuttur. Bu araştırmadan elde edilecek verilerin, günümüzde birçok ülkede uygulanan ve Türkiye' nin eğitim programlarında da yer verilmesi beklenen çokkültürlü eğitim uygulamaları hakkında ilgili literatüre katkı sağlayacağı umulmaktadır.

Araştırmanın Amacı: Bu araştırmanın amacı, her biri üniversite- lise veya ilköğretim düzeyinde öğretmenlik yapan ve eğitim programları ve öğretim bölümünde doktora eğitimini sürdüren kişilerin çokkültürlü eğitim algılarını ve değerlendirmelerini öğrenmektir.

Araştırmanın Yöntemi:

Nitel türdeki araştırmada durum çalışması deseni kullanılmıştır. Çalışmada "durum" olarak doktora öğrenimi gören öğretmen ve akademisyenlerin çokkültürlü eğitim algıları incelenmiştir. Araştırma, Türkiye'deki en büyük devlet üniversitelerinden birinin Eğitim Bilimleri bölümünde, doktora düzeyinde verilen Çokkültürlü Eğitim dersi kapsamında gerçekleştirilmiştir. Bu araştırmanın çalışma

grubunu, 2012-2013 eğitim- öğretim yılında bu dersi alan beşi öğretmen, dördü akademisyen olmak üzere dokuz doktora öğrencisi oluşturmuştur. Nitel araştırma yöntemlerinden yapılandırılmamış görüşme, gözlem, alan notları ve doküman analizi kullanılmıştır. Sonuçların daha objektif olması açısından haftalık tartışma yazıları (discussion board), sınav kağıtları ve gözlem formu kullanılmıştır. Bütün görüşmeler ses kaydı altına alınmış, sonra bilgisayar ortamına aktararak raporlaştırılmıştır. Nitel veriler *içerik analizi* yöntemiyle çözümlenerek, sonuçları raporlaştırılmıştır. Araştırmacı süreçte bulguları olduğu gibi ortaya koymaya çalışmış ve düşüncelerini sürece yansıtılmaya önem göstermiştir. Araştırmanın iç geçerliğini (inandırıcılığı) artırmak için veri çeşitliliğine gidilmiş, bulgular gözlem, görüşme ve doküman analizlerine dayalı olarak verilmiştir. Ses kaydına alınan soru ve cevaplar raporlandıktan sonra çalışma grubuna tekrar okutulmuş, katılımçıların teyidi alınmıştır. Görüşme ve gözlem bulguları ile doküman analizleri raporlandıktan sonra 2 alan uzmanına okutulmuş, onların görüşleri doğrultusunda düzenlemelerde bulunulmuştur. Araştırmanın güvenilirliğini (tutarlılığını) artırmak için ise bulguların sunumunda doğrudan alıntılara yer verilmiştir. Gözlem iki araştırmacı tarafından yürütülmüş, raporlama aşamasında bu gözlem bulguları karşılaştırılarak ortak yargılara varılmıştır. Görüşme ve gözlem bulguları ile doküman analizleri araştırmacılar tarafından okunarak olası kodlara ve temalara karar verilmiş, ardından bu tema ve kodların teyidi Nvivo 10 programı ile yapılmıştır.

Araştırmanın Bulguları: Araştırma bulguları analiz edildiğinde en çok vurgu yapılan kavramların eşitlik, demokrasi ve adalet olduğu görülmüş, çokkültürlü eğitimin bu kavramları güçlendireceği belirtilmiştir. Katılımcıların sorulara verdikleri cevaplar genel olarak analiz edildiğinde eşitlik kavramı üzerinde ciddi anlamda durulduğu ve çokkültürlü eğitimin eşitliği sağlama yönündeki önemi konusunda bir fikir birliğinin varlığı dikkati çekmiştir. Ayrıca çokkültürlü eğitimin demokratik bir toplum ve demokratik bir ortam sağlama konusundaki katkılarının sıkça belirtildiği, sosyal adaleti sağlama noktasındaki önemine sıkça vurgu yapıldığı ve bu özelliğinin çok önemsendiği görülmüştür. Bununla birlikte hoşgörü, saygı ve barış da sık tekrarlanan kavramlardan olmuştur. Yapılan gözlemler neticesinde ise ders yürütücüsünün yönlendirmesiyle her hafta gerekli makaleleri okuyarak ve çokkültürlü eğitimle ilgili verilen soruları tartışma yazısı şeklinde yazarak derse gelen öğrencilerin, sınıfta okuduklarından anladıklarını, çokkültürlü eğitim hakkındaki bazı endişelerini ve şaşırtdıkları noktaları, çokkültürlü eğitimin gerekçe ve hedeflerini, farklı kültürlerden gelen öğrenciler için faydalı olmasının nedenlerini ve Türk eğitim sistemi için faydalı olacak endişeleri tartıştıkları görülmüştür. Öğrenciler ilk hafta en çok çokkültürlü eğitimin ülkeyi böleceği endişesini taşıdıklarını ifade etmişlerdir. Ancak haftalar içinde yaptıkları okumalar doğrultusunda aslında çokkültürlü eğitimin birleştirici bir eğitim olduğunu ve daha eşit, demokratik ve adil bir toplum düzeni kurmayı sağladığını belirttikleri gözlenmiştir. Öğrencilerin tamamı çokkültürlü eğitimin gerekli ve önemli olduğu noktasında görüş birliğine varmış, farklı kültürlerden olan öğrencilerin asimile olmadan kendi değerlerini yaşatabilmeleri ve daha başarılı

olabilmeleri için de çokkültürlü eğitimin ayrı bir önem taşıdığını belirttikleri görülmüştür.

Araştırmanın Sonuçları ve Önerileri Bu araştırmada, literatürde belirtilenlere paralel olarak doktora öğrencilerinin en çok üzerinde durdukları kavramlar eşitlik, adalet ve demokrasi olmuştur. Katılımcılar, çokkültürlü eğitimin herkesin daha eşit fırsatlara ve eşit eğitim hakkında sahip olunmasına katkı sağlayacağını ifade etmişler, çokkültürlü eğitimin adaleti sağlama noktasındaki öneminden bahsetmişlerdir. Ayrıca farklı özelliklere ve kültürlere sahip kişilerin aynı toplumda eşit haklara sahip şekilde yaşayabilmeleri ve toplumun her kesimden insan için daha adil ve demokratik bir yapıya sahip olabilmesi için çokkültürlü eğitimin bir bakıma zorunlu olduğunu belirtmişlerdir. Çokkültürlü eğitim alanında Türkiye’de yapılmış olan çalışma sayısı sınırlı olmakla birlikte, tüm gelişmiş ve demokrasinin tam olarak uygulandığı ülkelerde olduğu gibi Türkiye’de de bu alanda yapılan araştırmaların artırılması, eğitimlerin verilmeye başlanması ve eğitim programlarının çokkültürlülüğü kapsayacak şekilde geliştirilmesi önerilmektedir.

Opinions of Secondary School Students on the Effects of Disciplinary Regulations and Practices on the Freedom of Thought and Expression*

Pelin TAŞKIN**

Suggested Citation:

Taşkın P. (2014). Opinions of Secondary School Students on the Effects of Disciplinary Regulations and Practices on the Freedom of Thought and Expression. *Eurasian Journal of Educational Research*, 57, 51-72 <http://dx.doi.org/10.14689/ejer.2014.57.5>

Abstract

Problem Statement: Disciplinary rules are necessary for students to benefit from education and training activities without any problems or shortcomings in the school environment. Governed by regulation in Turkey, these rules prescribe such penalties as reprimand, short-term suspension, changing of schools, or exclusion from formal education institutions. Conversely, various national and international legal arrangements define children's fundamental rights and freedoms. Students below the age of 18 are also entitled to fundamental rights and freedoms by reason of such legal arrangements. One of these arrangements recognises the freedom of thought and expression. The children's/students' freedom of thought and expression within the school environment, as well as the limitations thereof, are of a disputed nature. In light of such disputes, the problem addressed by the present study concerns the evaluation of the effects of disciplinary regulations and practices on students' freedom of thought and expression in line with students' own opinions.

Purpose of Study: The present study aims to identify how secondary school students consider the freedom of thought and expression within the context of disciplinary regulations and practices.

Methods: The study is a qualitative study designed and implemented in a phenomenological pattern. The sample of the study was composed, in line with the maximum diversity sampling technique, of 15 students from amongst all students enrolled at general public and private secondary

* The present study is based primarily on the PhD Thesis of Pelin Taşkın at Ankara University, supervised by Prof. Dr. Yasemin Karaman Kepenekci, called 'An Evaluation of Disciplinary Code on Secondary Education Students on the Basis of Fundamental Rights and Freedoms.'

**Dr. Ankara University Faculty of Educational Sciences, 06590, Ankara, Turkey, email:ptaskin@education.ankara.edu.tr

education institutions located in the central districts of the province of Ankara. Fifteen students were interviewed face-to-face in accordance with the semi-structured interview form. The interviews were recorded on paper and made subject to content analysis. The analysis categories (themes) were defined, in parallel with the objective of the study, as the evaluation of disciplinary regulations and practices applied to students in secondary schools within the dimension of the "Freedom of Thought and Expression". The subcategories were identified with the inductive method during the analyses.

Findings and Results: The opinions of students on the freedom of thought and expression were collected in the categories of "speech" and "participation". The opinions of students on such practices as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. for educational purposes in schools were assessed in the categories of "yes", "conditional yes", and "no". Their opinions concerning the conduct of such practices for political purposes were addressed in the categories of "yes" and "no". Finally, the opinions of students on the imposition of short-term suspension, expulsion, or exclusion from formal education institutions on students for such practices were assessed in the categories of "consider the penalty dysfunctional" and "consider the penalty both functional and dysfunctional". The students taking part in the study have a conceptual knowledge of the freedom in question and a perception of the same as the right to speak. The study concluded that students mostly included the practices of hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins or armbands, etc., for educational purposes in the scope of the freedom of thought and expression, but excluded the conduct of the same practices for political purposes from the scope of said freedom. Finally, the majority of participants found the imposition of disciplinary penalties on students for the conduct of such practices as dysfunctional.

Conclusions and Recommendations: In line with these results, it may be suggested that disciplinary rules be consistent with the age and maturity of students. Furthermore, objectivity and compliance with international norms and democratic society in the limitations concerning the freedom of thought and expression are of great significance.

Keywords: student rights, disciplinary regulations, freedom of thought, freedom of expression

The freedom of thought is defined as an individual's right to have any thought or opinion, to be free from any pressure by reason thereof, not to be forced to disclose such thoughts and opinions, and not to be reprimanded or blamed for the disclosure thereof. In the statements used by Teziç (1990, 33), the freedom of thought is the possibility for an individual to choose or prepare their potential responses to any emerging problems and to adapt their individual or social acts to such responses. The right to the freedom of thought is a right that protects not the thought, but rather the owner of the thought. The freedom of expression is, alternatively, defined as the freedom of expressing any belief, opinion, attitude, or emotion in an amicable

manner. This right protects the forms of thought and expression (Erdoğan, 2007, 19-21; Bezanson, 2005, 239).

Article 19 of the Universal Declaration of Human Rights (1948) recognises the freedom of thought and expression for everyone. This freedom enables the individual to be entitled to be free from any disturbance by reason of their thoughts and to research, obtain, and disseminate their knowledge and thoughts in any manner possible without being limited by national borders. Article 10 of the European Convention on Human Rights (1950) (ECHR) states that "[e]veryone has the right to freedom of expression" and the European Court of Human Rights (ECtHR) considers this statement as a mandatory basis for democratic societies and as one of the most prominent conditions for progress and development for everyone (Kaboğlu, 2000, 110). Pursuant to the aforementioned statement, the freedom of thought and expression may not be restricted unconditionally. Any limitation must be consistent with the aims of protecting the individual or public interests or maintaining or restoring the independence and impartiality of the judiciary. Article 13 of the United Nations Convention on the Rights of the Child (1989) (UNCRC) prescribes in the first paragraph that every child shall have the right to express their thoughts freely; this right may be exercised, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other media of the child's choice. However, the second paragraph of Article 13 states that the exercise of this right may be subject to certain restrictions, but these shall only be such as are provided by law and are necessary for respect of the rights or reputations of others; or for the protection of national security or of public order or of public health or morals. Moreover, Article 14 of UNCRC prescribes an obligation for state parties to respect the right of the child to freedom of thought, conscience, and religion.

In Turkish law, the Constitution of 1982 regulates the freedom of thought and expression in Articles 25 and 26 and does not provide a distinction between adults, children, or students. The exercise of the said freedom may be restricted for the purposes of national security; public and safety; safeguarding the basic characteristics of the Republic and the indivisible integrity of the State with its territory and nation; preventing crime; punishing offenders; withholding information duly classified as a state secret; protecting the reputation, rights, and private or family life of others; protecting professional secrets as prescribed by laws or ensuring the proper functioning of the judiciary.

The child's/student's freedom of thought and expression has been discussed in literature. Some authors argue that children/students cannot be entitled to the freedom of thought and expression. As an example, Brighouse (2002, 51) asserts that children grow in a family environment where they are influenced greatly by their parents, and most of their knowledge comes from this impact. Although it is important for the development of children to be allowed to express themselves, it must be known that such expression does not fully reflect their own personalities and comments. Similarly, according to Etzioni (2004, 4), principles must be defined for the interpretation of the freedom of thought and expression, and such definitions must observe the best interests of the child. Etzioni's approach is essentially based on the child as a creature in need of protection and the importance of protecting them against any potential harm. In contrast, liberals also admit that the child must be protected against certain types of freedom of thought and expression. According to

the liberal opinion held by Macleod (2004, 57), children have a specific, advanced interest based on the concept of moral personality. Every child must develop and utilise their moral power, which shapes the distinct and independent moral personalities of individuals. This argument leads to the emergence of the child's interest in obtaining information and developing independent opinions and thoughts. By reason of these interests, children are entitled to the freedom of thought and expression before adulthood.

The literature does not present a consensus concerning whether or not students are entitled to the freedom of thought and expression within the school environment. Especially in the Anglo-Saxon legal system, the scope and limits of freedom are defined in line with judicial decisions. However, there is no judicial decision in Turkey with respect to students' freedom of thought and expression. Instead, certain provisions of the Regulation on Secondary Education Institutions under the Ministry of National Education (e.g., 164/(1)/(j) and 164/(2)/(g)) dated 07.09.2013 includes restrictive arrangements on this matter. In parallel with the restrictive arrangements in the aforementioned regulation, Turkey unfortunately witnesses incidents that give rise to the impression that students' freedom of thought and expression is not addressed with a warm attitude. One of these incidents occurred in June 2014 when a best student, İştan Önder, received a reprimand after using the following statements in his graduation speech: "They have taken their right to life away. It is impossible for anyone not to think of Berkin Elvan or Ali İsmail Korkmaz. You must know that Ali İsmail Korkmaz, Berkin Elvan, and others will always be sitting on these desks." His best student status was taken away due to the penalty in question; consequently, he did not benefit from the best students' quota in the university entrance exam. Together with Eğitim-Sen Union, İştan Önder's family filed a suit at Kocaeli Administrative Court for the restitution of his best student status. The 2nd Administrative Court of İzmit granted a stay of execution on the grounds that such a penalty could not be imposed. Nevertheless, the case has not yet been concluded (Hürriyet, Cumhuriyet, 2014).

The present study aims to identify how secondary school students consider the freedom of thought and expression within the context of disciplinary regulations and practices. In line with this objective, the study sought their opinions on the effects of disciplinary regulations and practices applied to secondary school students concerning their freedom of thought and expression, one of the fundamental rights and freedoms of children/students.

Method

Research Design

The present study aimed to identify how secondary school students consider the freedom of thought and expression within the context of disciplinary regulations and practices. Therefore, the study was designed and implemented in a phenomenological pattern that aimed to research the phenomena encountered in the form of incidents, experiences, perceptions, trends, concepts, and situations (Yıldırım and Şimşek, 2005, 72).

Target Universe and Study Sample

The study's target universe was composed of students enrolled at general public and private secondary education institutions affiliated with the Ministry of National

Education and located in the nine central districts of the province of Ankara, namely Altındağ, Çankaya, Etimesgut, Gölbaşı, Keçiören, Mamak, Pursaklar, Sincan, and Yenimahalle in the 2012-2013 Educational Year. The sample of the study was composed, in line with the maximum diversity sampling technique, of 15 students from amongst all students enrolled at general public and private secondary education institutions located in the central districts of the province of Ankara.

There are general public secondary education institutions in all central districts of the province of Ankara. However, there are no private secondary schools in Mamak and Sincan. The private secondary education institutions located in Gölbaşı and Keçiören did not permit interviews or failed to respond to the request for such permission. Therefore, the study was implemented at private secondary education institutions located only in Çankaya, Etimesgut, Pursaklar, and Yenimahalle. The private secondary education institution located in Altındağ was under the same ownership as the private secondary education institution located in Pursaklar and was therefore excluded from the study. As the number of general private secondary schools was too low when compared to the number of general public secondary schools, two more private schools from the district of Çankaya were also added to the scope of the study. Information regarding the codes, sexes, and localities of the participants is given in Table 1.

Table 1.

Expansion of the Codes Corresponding to Students in the Study Sample

SK1k	public	1. row	female	Keçiören.
SK2k	public	2. row	female	Çankaya
SK3k	public	3. row	female	Gölbaşı
SK4k	public	4. row	female	Etimesgut
SK5k	public	5. row	female	Yenimahalle
SK6e	public	6. row	male	Altındağ
SK7e	public	7. row	male	Mamak
SK8k	public	8. row	female	Sincan
SK9k	public	9. row	female	Pursaklar
SÖ1k	private	1. row	female,	Çankaya
SÖ2k	private	2. row	female	Yenimahalle
SÖ3e	private	3. row	male	Etimesgut
SÖ4k	private	4. row	female	Pursaklar
SÖ5e	private	5. row	male	Gölbaşı
SÖ6e	private	6. row	male	Çankaya

Data Collection

For the purposes of collecting data in line with the objective of the present study, a focus group meeting was held with a group of 7 teachers and administrators on 11.06.2012 with the aim of developing the statements/questions to be included in the

student interview forms. The interview forms were prepared on the basis of the results of this focus group meeting and the "Reward and Discipline Regulation dated 19.01.2007 and No. 26408 for Secondary Education Institutions under the Ministry of National Education", which was in effect in this term (the second term of the 2012-2013 school year). Then, the interviews were submitted to an expert for review.¹ The interview forms constitute the data collection tools for qualitative research (Punch, 2005, 165). Interview forms allow for the systematic collection of information from a group of individuals with the same questions (Patton, 1987 trans. Kümbetoğlu, 2005, 75). The benefit of using interview forms lies in the most effective utilisation of the generally limited interview duration (Kümbetoğlu, 2005, 75). In this study, semi-structured interview forms were preferred for data collection, as this method provides in-depth information regarding the subject at hand. The interview forms were finalised in line with the expert's opinions.

Permission was obtained from the Provincial Directorate of National Education (dated 12.04.2013 and No. 14588481/605.99/573725) for the application of the "Interview Form on Students' Opinions on the Reward and Discipline Regulation for Secondary Education Institutions under MoNE and its Implementation within the Context of Students' Fundamental Rights and Freedoms" to secondary school students.

Fifteen students were interviewed face-to-face in accordance with the semi-structured interview forms. The interviews were conducted by the author. All participants were asked for permission for the sound recording of interviews; interviews with one student from a public secondary education institution and two students from private secondary education institutions were recorded on paper and the rest of the interviews were sound recorded. The majority of the interviews were conducted in the respective school environment; however, interviews with two students from private secondary education institutions were held in localities of their choosing outside their respective schools due to the start of the summer holiday.

Data Analysis

The sound recordings taken during interviews were transcribed. The transcribed or noted opinions of the participants were compiled into written text in the computer medium. Interviews thus compiled in written text were made subject to content analysis. Fröh defined content analysis as "an empirical method providing a systematic and objective description of contextual formal characteristics of statements" (Fröh, 2001 trans. Gökçe, 2006, 17). "Content analysis, as a method, aims to obtain certain findings on some dimensions and cross-sections of non-existent, or unknown social reality on the basis of the quantitative and qualitative dimensions of existing texts" (Gökçe, 2006, 20). In line with this aim, the first step to content

¹ Prof. Dr. Ali Balcı, Prof. Dr. İnyet Aydın, Prof. Dr. Nejla Kurul, Prof. Dr. Işıl Ünal, Prof. Dr. Meral Uysal, Prof. Dr. Ezel Tavşancıl, Prof. Dr. Hasan Hüseyin Aksoy Assoc. Prof. Dr. Şakir Çınkır, Ass. Prof. Dr. Nihan Demirkasımoğlu, Res. Ass. Dr. Saadet Kuru Çetin, Res. Ass. Dr. Fatih Kezer.

analysis is defining an analysis category (main category) and subcategories (Strauss and Corbin, 1990 trans. Yıldırım and Şimşek, 2005, 227).

The analysis categories (themes) were defined, in parallel with the objective of the study, as the evaluation of disciplinary regulations and practices applied to students in secondary schools within the dimension of the "Freedom of Thought and Expression". Although this category was formed at the start of the research through the inductive method, the subcategories were identified during the analyses through the deductive method (Gökçe, 2006, 20). In accordance with the study problem, the first theme was associated with the first sub-problem, while the second, third, and fourth themes were associated with the second sub-problem. The author worked with his thesis supervisor in the codification of the dataset and reached a common codification by discussing the similarities and differences of all codifications and conceptualisations regarding the data.

The frequency and percentage values of the breakdown of participants' opinions were calculated and this breakdown was represented in tables. The interpretation of the participants' opinions also utilised the sentences/statements used by the participants themselves.

The analysis results obtained from the research were also codified by two experts² with experience in quantitative research and previous academic research on fields relevant to the thesis subject with a view to testing the consensus. The interview transcripts of the study were codified primarily by the author. In addition to the author, two independent experts repeated these codifications. The results of codifications undertaken by different experts were compared through the formula developed by Miles and Huberman (1994, 278).

As a result of the codifications, the reliability degree was determined to be 79% with the first expert and 76% with the second expert. Houten and Hall (1983, 27) considers a reliability degree of 70% sufficient. Therefore, it was concluded that analyses were performed with a sufficient degree of reliability.

Results

Students' Opinions on Freedom of Thought and Expression

The students were asked to define the freedom of thought and expression and requested to explain this freedom. The students' opinions are given in Table 2.

² Experts: Ass. Prof. Dr. Funda Nayır and S. İpek Aksoy Gülşen.

Table 2.*Breakdown of Participants' Opinions on the Freedom of Thought and Expression*

Study Question	Sample Statements	Themes	f
What do you think is Students' Freedom of Thought and Expression?	<i>It is to express one's own opinions anywhere. (SÖ3k; SÖ4k)</i>		
	<i>It is to express what one knows and wants to say with ease and openness. (SK3k; SK7e; SÖ1k; SÖ2k)</i>	Speech	11
	<i>You can express your own thoughts without breaching the line of respect, but you shouldn't overlook the fact that you are talking to a teacher, to an older person. You can thus express yourself and explain your thoughts. (SK4k; SK5e; SÖ8k; SÖ9k)</i>		
	<i>I think, it means that I am free to draw pictures in the classroom. ... (SÖ6e)</i>		
	<i>It is to express one's thoughts and opinions freely. It is the ability to express your opinions and suggestions to higher bodies and for them to take these into consideration. ... (SK1k; SK2k; SK6e, SÖ5e)</i>	Participation	4

As can be seen in Table 2, the opinions of 11 students (6 public, 5 private) were addressed in the subcategory of "the right to speak", whereas the opinions of 4 students (3 public, 1 private) were included in the subcategory of "right to participate".

The students' opinions addressed in the subcategory of the right to speak were observed to define the freedom of thought and expression as the right to express one's own thoughts (SÖ3e, SÖ4k). Moreover, some students stated that the freedom of thought and expression is the right to speak freely and without any pressure (SK3k, SK7e, SÖ1k; SÖ2k). Some students also defined the freedom of thought and expression as the right to speak, but added that this freedom is restricted by a line of respect (SK4k, SK5k, SK8k, SK9k). In contrast, the majority of the students were seen to perceive expression as verbal expression. Nevertheless, SÖ6e stands out from other students in that they do not perceive expression only as verbal expression, but state that expression is also possible through painting. In addition, certain students stated that they perceived the freedom of thought and expression as their ability to express their own emotions and thoughts regarding any situation to adults and especially to school teachers and to be included in decision-making processes of relevance for themselves (SK1k, SK2k, SK6e, SÖ5e).

Opinions of Students on the Inclusion of such Practices as Hanging Banners, Distributing Brochures and Leaflets, Printing Newspapers or Magazines, or Wearing Pins And

Armbands, etc. for Educational Purposes within the Scope of the Freedom of Thought and Expression

The opinions of students on the question as to whether such practices as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. for educational purposes can be addressed within the scope of the freedom of thought and expression are given below (Table 3).

Table 3.

Breakdown of Students' Opinions on the Inclusion of such Practices as Hanging Banners, Distributing Brochures and Leaflets, Printing Newspapers or Magazines, or Wearing Pins and Armbands, etc. for Educational Purposes within the Scope of the Freedom of Thought and Expression

Study Question	Sample Statements	Themes	f
Do you think such practices as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. for educational purposes can be addressed within the scope of the freedom of thought and expression? Why?	<i>I think they can. In fact, students' opinions must be known to all and if there is going to be a change in the education system, it should be affected in line with these opinions. I believe that the student can express their opinions on such matters. (SK1k)</i>		
	<i>Of course it can. This is expressing an opinion, showing an opinion. Regardless of your opinion, you have the right to express your opinion. To that end, we can also hang banners. Because you have peers at school, who understand you and value your opinions. So, I believe hanging a banner to disseminate your opinions to them is quite right. (SÖ5e) (SK2k; SK7e; SK9k; SÖ1k; SÖ3e; SÖ4k; SÖ6e)</i>	Yes	9
	<i>They should do that. To give us information on something, they can use different ways to achieve higher goals in education. With different elements, for example without any prohibitions, they can do something properly within the framework of the law. (SK3k)</i>		
	<i>If it is indeed educational. Our principal keeps warning us already. If there is nothing, no behaviour that can have a bad effect on us... but I think it shouldn't be done without the approval of the principal. (SK4k)</i>		
	<i>Last year, we had a collective mourning for a martyr. Then, they distributed red ribbons to everyone. All students, including us, and all teachers walked around with those. I think it should be done if it does not disturb the peace or affect people adversely. (SK8k)</i>		
	<i>If educational, it can be. But if a student is going to do this, they shouldn't go over the disciplinary limit. (SÖ2k)</i>	Conditional Yes	4
<i>There can be opposite opinions. And that will not be good. It shouldn't be in the school, but can be done outside of the school. (SK5k)</i>			
<i>I think these are all wrong behaviours. We have education at school and we haven't reached 18 yet and we come to the school as students. So once we are 18, we can do whatever we want outside, but it will not be appropriate for us to do this within the school limits. (SK6e)</i>	No	2	

As can be seen in Table 3, the opinions of participating students were divided into the subcategories of "yes", "conditional yes" and "no" and assessed accordingly. In line with this assessment, the opinions of 10 students (5 public, 5 private) were addressed in the subcategory of "yes", while the opinions of 3 students (2 public, 1 private) were assessed in the subcategory of "conditional yes" and the opinions of 2 students from public schools were in the subcategory of "no".

The students whose opinions were addressed in the subcategory of "yes" saw such practices as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. as a way to express and disclose their opinions as communication elements and considered them within the scope of the aforementioned freedom. These methods can be used to change the education system according to SK1k and to share opinions with peers for SÖ5e. The students' opinions addressed in the subcategory of "conditional yes" admit that the students can actually conduct the said activities for educational purposes, but add that such activities should not lead to negative consequences in short term. The negative consequences implied here were defined as in contradiction to laws (SK3k), misdirection (SK4k), disturbance of peace/affecting people (negatively) (SK8k), and disturbance/breach of discipline (SÖ2k). Considering the opinions assessed in the subcategory of "no", these students can be stated to consider such acts as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. as problematic regardless of their purpose.

Breakdown of Students' Opinions on the Inclusion of such Practices as Hanging Banners, Distributing Brochures and Leaflets, Printing Newspapers or Magazines, or Wearing Pins And Armbands, etc. for Political Purposes within the Scope of the Freedom of Thought and Expression

The opinions of students on the question as to whether such practices as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. for political purposes can be addressed within the scope of the freedom of thought and expression are given below (Table 4).

Table 4.

Breakdown of Students' Opinions on the Inclusion of such Practices as Hanging Banners, Distributing Brochures and Leaflets, Printing Newspapers or Magazines, or Wearing Pins and Armbands, etc. for Political Purposes within the Scope of the Freedom of Thought and Expression

Study Question	Sample Statements	Themes	f
Do you think such practices as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. for political purposes can be addressed within the scope of the freedom of thought and expression? Why?	... I think these should be allowed, but there have to be certain lines, red lines. This should not turn into a left or right matter. Whenever I write something about workers, I am immediately blacklisted, "oh, he is a leftist"... this shouldn't be the case. This is actually an educational activity. My parents are both workers. I believe that when I do this, it shouldn't be considered as political. (SK7e)	Yes	4
	I think they should be able to do this for political purposes, as well. Because in fact, this is freedom of expression. They should be able to express their opinions without someone stopping them and forcing them to adopt a certain opinion, without targeting someone else's right. (SÖ5e) (SK2k, SÖ1k)		
	No, it cannot. Then, there would be an environment of arguments, more fights. So it would be more difficult to restore order. (SK3k) (SK1k, SK5k, SÖ3e, SÖ4k)		
	If it is for political purposes, it is a crime in itself. Because an individual of that age cannot have anything to do with politics. (SK4k)	No	11
	As we are below the age of 18, we cannot (act) like this. We perceive political behaviours in our own framework and there can be differences of political opinion at school, too. So it will have a negative effect if done within the school. (SK6e) (SK8k, SK9k, SÖ2k, SÖ6e)		

As can be seen in Table 4, the opinions of participating students were divided into the subcategories of "yes" and "no" and assessed accordingly. In line with this assessment, the opinions of 4 students (2 public, 2 private) were addressed in the subcategory of "yes", while the opinions of 11 students (7 public, 4 private) were assessed in the subcategory of "no". The students' opinions in the category of "yes" considered the said acts within the scope of the freedom of thought and expression, even if they were realised for political purposes. In the study, the definition of possible political purposes was left to the students to consider. Thus, SK7e emphasised that matters that appear to be political to certain people may actually have an educational aspect with their words regarding the educational and non-political nature of statements on workers. Conversely, SÖ5e states that the consideration of the acts in question within the scope of the freedom of thought and expression are not problematic unless someone forcefully imposes their opinions on someone else. Considering the opinions in the subcategory of "no", we see that

students do not consider the realisation of such acts for political purposes within the scope of the freedom of thought and expression. The justifications put forth by the students are young age, potential for arguments/conflicts, criminal aspect/prohibition, differences of opinion, going beyond the purposes of the school, and polarisation.

Opinions of Participants on the Imposition of Short-Term Suspension, Expulsion, or Exclusion from Formal Education Institutions as Penalties due to such Acts as Hanging Banners, Distributing Brochures and Leaflets, Printing Newspapers or Magazines, or Wearing Pins and Armbands, etc. at School

The opinions of participants on the imposition of short-term suspension, expulsion, or exclusion from formal education institutions as penalties due to such acts as hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins and armbands, etc. at school are addressed below (Table5).

Table 5.
Breakdown of Opinions of Participants on the Imposition of Disciplinary Penalties due to such Acts as Hanging Banners, Distributing Brochures and Leaflets, Printing Newspapers or Magazines, or Wearing Pins and Armbands, etc. at School

Study Question	Sample Statements	Themes	f
How do you feel about the imposition of short-term suspension, expulsion, or exclusion from formal education institutions on a student due to such acts?	<i>I don't think that such penalties will be very effective on students. Because when you impose a penalty, the problem is not solved at its root. Instead, the student can be warned not to repeat the same act or if he doesn't know that act is wrong, such acts should be explained to students at the beginning so that they do not act in that wrong way. So they should exercise the freedom of thought in any way except for political purposes. (SK1k) (SK8k, SK2k)</i>	Consider penalties dysfunctional	11
	<i>If we are living in a free place, the student should be able to express their emotions and opinions. He should not be penalised for this. If he did something wrong, he should be reprimanded. (SK3k) (SÖ3e, SÖ5e, SÖ6e)</i>		
	<i>I am absolutely against this. No student should be excluded from his environment due to his opinions. In fact, this is where the freedom of thought is restricted. (SK9k) (SK5k, SK7e, SÖ1k)</i>	Consider penalties both functional and dysfunctional	4
	<i>Actually, a reprimand... may be normal. But I think he doesn't have to be expelled or sent to an open high school. He does it once. ...But he won't do it again after a reprimand. (SÖ2k)</i>		
	<i>... I think suspension should not be an option, but firstly a reprimand should be issued. If the reprimand does not work, then other penalties can be considered. I think, suspension should be the last resort. (SK4k) (SK6e, SÖ4k)</i>		

As can be seen in Table 5, the students' opinions were assessed within the subcategories of "consider penalties dysfunctional" and "consider penalties both functional and dysfunctional". In line with this assessment, the opinions of 11 students (7 public, 4 private) were addressed in the subcategory of "consider penalties dysfunctional", while the opinions of 4 students (2 public, 2 private) were assessed in the subcategory of "consider penalties both functional and dysfunctional".

With respect to the opinions addressed in the subcategory of "consider penalties dysfunctional", students do not consider the imposition of penalties on students for exercising their freedom of thought and expression and argue that the freedom of thought and expression cannot be fully exercised due to such penalties. In addition, the students whose opinions were assessed in this subcategory think that the imposition of penalties due to the exercise of the freedom of thought and expression has a negative effect on their freedoms, but it is not appropriate for them to exercise this freedom in such a manner as to reflect political elements. In terms of the opinions addressed in the subcategory of "consider penalties both functional and dysfunctional", the students do not consider the imposition of penalties for the said acts as problematic, but are of the opinion that the current penalties are too heavy and less substantial penalties should be imposed, such as warnings and reprimands.

Discussion and Conclusion

The present study explained how the students' freedom of thought and expression is regulated in national and international legislation, the current situation in various legal systems, and how students perceive the freedom in question. According to the study results, the students taking part in the study have a conceptual knowledge of the freedom in question and a perception of the same as the right to speak. The study concluded that students mostly included the practices of hanging banners, distributing brochures and leaflets, printing newspapers or magazines, or wearing pins or armbands, etc., for educational purposes in the scope of the freedom of thought and expression, but excluded the conduct of the same practices for political purposes from the scope of said freedom. Finally, the majority of participants found the imposition of disciplinary penalties on students for the conduct of such practices as dysfunctional (Lobel, Heckel and Avenarius, 2000, 548)

Whether or not the students have the freedom of thought and expression in schools has been discussed in legal literature. Considering these discussions, there are two approaches in the Anglo-Saxon law: i.e., restrictive and liberal. The restrictive approach argues that student freedom of thought and expression at school is not equal to adult freedom of thought and freedom. In fact, the school system pursues the sole goal of educating students. Therefore, any expression may be restricted by the school administration if it contradicts the goals of the school or the pedagogical duties of schools (Landwehr, 2000, 393). Alternatively, the liberal approach holds schools as institutions where students find openings to express their

opinions by taking the floor in classrooms or through different means in school environments (e.g., specialised community activities, school newspaper, etc.). As with the other members of the society, students feel entitled to the freedom of thought and expression and it is admitted that these freedoms are maintained also within the physical borders of a school. Thus, the school administration may not impose restrictions that will eliminate students' freedom of expression, even if it does not agree with some of the students' opinions and/or the methods they use to express the same. The administration may restrict the freedom of thought and expression to attain certain goals – those consistent with the requirements of a democratic society and pursue the realisation of the common interests of the public - and the proportionality of the restrictions applied; the goals pursued must be observed in this process (Imber and Geel, 2005, 71).

Similarly to the liberal approach in the Anglo-Saxon system, the German Law (pursuant to Article 5(1) of the Basic Law for the Federal Republic of Germany, *Grundgesetz*, GG) admits that students are entitled to the freedom of thought and expression in any relationship they establish with the school. German Law prescribes training students with the skills needed to have a perspective and advocate various theses on any matter expressly as one of the duties of schools and considers any school that hinders students' freedom of thought and personal development to be in contradiction with its educational duties. Nevertheless, the freedom of thought and expression granted to students is not free from restrictions; the restrictions specified in provisions in general laws are aimed at protecting the youth, human reputation, and dignity (Article 5(2) of the Constitution). The general laws not only define the duties of schools, but also the restrictions to the freedom of thought and expression. The general laws are interpreted in light of fundamental rights. The duties of schools are specified in general laws, as well as school laws where educational duties are concretised on the basis of their designations in national and state laws. However, students are not allowed to talk about matters other than subjects at hand or to talk during classes within the scope of the freedom of thought and expression. The teacher is in charge of avoiding any such disturbances.

The general objectives of education in Turkey are stipulated in Article 2 of the Basic Education Law No. 1739. According to paragraph 2, one of the objectives to be pursued by education is to “train students as constructive, creative and efficient individuals with personalities and characters developed in a balanced and healthy manner in terms of physical, mental, moral, spiritual and emotional aspects; equipped with the power of free and scientific thinking and a wide outlook of the world; approaching human rights with respect and valuing personalities and entrepreneurship; and having a sense of responsibility towards the society”. Setting off from this objective, it is possible to assume that students are entitled to the freedom of thought and expression in order for them to have “the power of free and scientific thinking and a wide outlook of the world”. Conversely, the Constitution of 1982 strengthens this assumption in that it regulates the freedom of thought and expression without a distinction between adults and children or students in Articles

25 and 26. The freedom in question may only be restricted for the purposes defined in the Constitution and in general laws.

Judicial decisions regarding students' freedom of thought and expression in Turkey mostly concern students enrolled in higher education institutions. However, these students are excluded from the scope of the present study. In any case, it must be noted here that the 8th Chamber of the Council of State cancelled the amendment to the "Student Discipline Regulation for Higher Education Institutions" that had added the act of "distributing leaflets" to the list of disciplinary offences to be reprimanded on the grounds that such an amendment constituted a breach of the freedom of thought and expression (Sondurakhaber, 2014).

In contrast, relevant judicial decisions can be seen in the USA, where educational law is quite advanced. The United States Supreme Court handled the presence or absence of students' freedom of expression and the potential situations giving rise to this freedom for the first time with its decision in *Tinker v. Des Moines Independent Community School District* in 1969. This decision resulted in the rule known as the *Tinker Test*. According to this rule, the school administration may only prohibit an expressed statement if such statement may cause a problem, lead to a material or substantial fault in the functioning of the school, or breach someone else's rights. The most important benefit of this test is that the school administration cannot penalise or prohibit any opinion merely on the grounds that it does not agree with the expression or solely on the grounds of a general fear of potential fault in the school system. Nevertheless, reasonable restrictions may be imposed by the school management with relation to the time, place, and manner of expression on the condition that such restriction(s) is required for the pursuit of educational goals. As an example, it is reasonable to prohibit students from talking politics during a maths class, but it is not reasonable to prohibit them from talking politics during lunch, i.e. at times where they can engage in free discussions regarding their own choices. However, for certain there are exceptional conditions.

Contrary to Turkey, the political opinions of students are welcomed within schools in Europe and the USA. As an example, according to Swiss and Austrian law, students can express their political opinions, publish them in the school newspaper, and form political organisations and assemblies even within schools. However, these rights may be restricted by lawmakers for educational purposes in cases where, for instance, militant student groups or political propaganda are observed. Richter (1990, 132) suggests that practitioners of educational law and school administrators find a common ground between the political impartiality of schools and political rights of students. Similarly, German Law admits the possibility of wearing pins and badges to express opinions and prescribes that the right to express opinions in this manner is under legal protection. Students can defend their opinions on social criticism and social morality in an excessive and influential manner (*recht auf jugendlichen Überschwang*). However, students cannot jeopardise the performance of educational duties and, specifically, cannot disturb the peace in schools. Yet, a student does not exceed the limits of the freedom of expression by merely wearing a badge (Avenarius, 2001, 105). Conversely, in the *Texas v. Johanson* case, the Supreme Court

ruled that students' clothing, hairstyle, rosaries, or badges, etc. can be considered within the scope of the freedom of thought and expression, whereas words of obscenity, threat, hostility, or insult cannot be addressed within the scope of the freedom of thought and expression and cannot benefit from legal protection (Imber and Geel, 2005, 51).

Suggestions

In order for students' freedom of thought and expression to be under legal guarantees, students must be primarily given information about the content and limits of the freedom in question. With respect to the specific statements of students that can be addressed within the scope of the freedom of thought and expression, the definition of such statements through legal arrangements is of great importance, as the Turkish legal system lacks a sufficient body of case-law in this field. The consistency of the legal arrangement in question with the age and maturity of students is also an important aspect. When we consider the rules in effect for the penalisation of disturbing behaviours of students, we see that students are addressed almost as adults. In fact, disciplinary rules for students must not be structured with "concepts and institutions from the adults' world" (Ümit Atılgan, 2007, 289).

Then again, the exercise of the freedom of thought and expression by students without any prior permission from an institution is not subject to penalties in the legal systems of foreign countries. Indeed, such acts are penalised only if they have exceeded certain limits predefined by law or following an assessment to be conducted in line with the legal facts of individual cases. However, students in Turkey can exercise their freedom of thought and expression only by obtaining permission from the school management or within the scope of specialised communities. The permission in question must be eliminated as it constitutes an obstacle to the effective exercise of the freedom of thought and expression and is replaced by criteria that appropriately restrict this freedom in an objective manner that is consistent with international norms and the requirements of a democratic society.

Finally, it is important to raise awareness among students in their exercise and protection of fundamental rights and freedoms with respect to the fact that they will be entitled to vote at the age of 18 and may even be involved in the executive activities of the state as public personnel or in legislative activities as elected officials in the national assembly.

References

- Avenarius H. (2001). *Einführung in das schulrecht [Introduction to school law]*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Basic Law for the Federal Republic of Germany, *Grundgesetz*, GG, 23 May., 1949, art.5. Retrieved April 20, 2014 from <https://www.btg-bestellservice.de/pdf/80201000.pdf>
- Best student status in Danger (2014, 18 June) Hürriyet. Retrieved June 19, 2014 from <http://www.hurriyet.com.tr/gundem/26627592.asp>
- Bezanson P. B. (1995). Institutional Speech, *Iwo Law Review*, 80, 735-824.
- Brighthouse H. (2002). What Rights (if any) Do Children Have?, In D. Archard and C. M. Macleod (Eds) *The moral and political status of children* (pp. 31-52). Oxford University Press: Oxford.
- Constitution of the Republic of Turkey, 1982, no. 2708, P.2, C.2, art.25, 26. Retrieved April 20, 2014 from http://global.tbmm.gov.tr/docs/constitution_en.pdf
- Erdoğan M. (2007). İfade Özgürlüğü ve Sınırları [Freedom of Thoughts and Limitations]. T. Bora (Eds), *İfade özgürlüğü ilkeler ve Türkiye [Freedom of thoughts: principles and Turkey]*.İstanbul: İletişim Yayınları.
- European Convention on Human Right, 4 Dec., 1950. Retrieved April 20, 2014 from http://www.echr.coe.int/Documents/Convention_ENG.pdf
- Etzioni A. (2004). On Protecting Children From Speech, *Chicago-Kent Law Review*, 79(3), 3-53.
- Good news for Işitan Önder (2014, 17 July) Cumhuriyet. Retrieved July 18, 2014 from http://www.cumhuriyet.com.tr/haber/turkiye/95301/Okul_birinciligi_eli_nden_alinan_Isitan_Onder_e_ iyi_haber.html
- Gökçe O. (2006). *İçerik analizi kuramsal ve pratik bilgiler [Content analysis: theoretical and practical information]*. Ankara: Siyasal Kitabevi.
- Houten V. R., Hall R.V. (1983). *The measurement of behaviour: behaviour modification*, USA.
- Imber M., Geel T. V. (2005). *A Teacher's guide to education law*. Roudledge: United States of America.
- Kaboğlu İ.Ö. (2000). Düşünce özgürlüğü [Freedom of thoughts] K. Tankuter (eds), *İnsan hakları [Human rights]*. İstanbul: YKY Yayınları.
- Kümbetoğlu B. (2005). *Sosyolojide ve antropolojide niteliksel yöntem ve araştırma [Qualitative method and research in sociology and anthropology]* Ankara: Bağlam Yayıncılık.

- Landwehr, A. (2000). A Student's Right to Freedom of Speech in Light of *Henerey v. City of St. Charles School District*, *Journal of Law and Education*, 29, 393-397.
- Lobel H.C., Heckel H., Avenarius H. (2000). *Schulrechtskunde [Information of Right to Education]*. Neuwied: Luchterhand.
- Macleod C. M. (2004). A Liberal Theory of Freedom of Expression For Children, *Chicago-Kent Law Review*, 79 (3), 55-82.
- Miles M. B., Huberman A. M. (1994). *Qualitative data analysis*, USA:Sage Publications.
- Punch K. F. (2005). *Sosyal arařtırmalara giriş, nicel ve nitel yaklaşımlar [Introduction to social researches, quantitative and qualitative approaches]* (Translated by D. Bayrak, H.B. Arslan, Z. Akyüz), Ankara: Siyasal Kitabevi.
- Regulation of Secondary Education Institutions, published in *official journal* on 07.09.2013, no. 28758, art. 164. Retrieved April 21, 2014 from http://mevzuat.meb.gov.tr/html/ortaogrkurumyon_0/ortaogrkurumyon_1.html
- Reward and Discipline Regulation, published in *official journal* on 19.01.2007, no. 26408, art. 12. Retrieved April 21, 2014 from http://mebk12.meb.gov.tr/meb_iys_dosyalar/16/01/964259/dosyalar/2013_03/08023339_odulvedisiplinyonetmeligi.pdf
- Richter I. (1990). West Germany, Switzerland and Austria, In I. Birch, I. Richter (Eds.), *Comparative School Law* (p.119-139).
- Sondurak (2014). Retrieved from <http://www.sondurakhaber.com/haber-10000-ogrenci-disiplin-yonetmeliginde-yapilan-degisikliklerin-yurutmesi-durduruldu.html>
- Student Discipline Regulation for Higher Education Institutions, published in *official journal* on 18.08.2012, no. 28388. Retrieved April 21, 2014 from http://www.yok.gov.tr/web/guest/icerik//journal_content/56_INSTAN CE_rEHF8BIsfYRx/10279/17960
- Texas v. Johnson, 491 U.S. 397 (1989).
- Teziç E. (1990). Türkiye'de siyasal düşünce ve örgütlenme özgürlüğü [Freedom of political thoughts and organization in Turkey]. *Anayasa Yargısı Dergisi*, 7, 29-46.
- Tinker v. Des Moines Independent Community School District, 393 U.S. 503 (1969).
- United Nations (1948/2009). Universal declaration of human rights. Retrieved April 20, 2014 from <http://www.un.org/en/documents/udhr/inde>

United Nations Convention on the Rights of the Child, 20 Nov., 1989. Retrieved April 20, 2014 from <http://www.ohchr.org/en/professionalinterest/pages/crc.aspx>

Ümit Atılğan E. (2007). *Mekandan imkana: çocuk suçluluğunun habitusu ceza ehliyeti ilişkisi* [From urban to opportunity: the habitus of juvenile delinquency]. Ankara: Ankara Barosu.

Yıldırım A., Şimşek H. (2005). *Sosyal bilimlerde nicel araştırma yöntemleri* [Methods of Quantitative Research in Social Sciences]. Ankara: Seçkin Yayınları.

Orta öğretim Öğrencilerinin Disiplin Düzenleme ve Uygulamalarının Düşünce ve İfade Özgürlüğüne Etkisi Hakkındaki Görüşleri

Atıf:

Taşkın P. (2014). Opinions of Secondary School Students on the Effects of Disciplinary Regulations and Practices on the Freedom of Thought and Expression. *Eurasian Journal of Educational Research*, 57, 51-72 <http://dx.doi.org/10.14689/ejer.2014.57.5>

Özet

Problem Durumu: Okullarda öğrencilerin sorunsuz ve eksiksiz biçimde eğitim öğretim faaliyetlerinden yararlanabilmeleri için disiplin kurallarına ihtiyaç duyulması kaçınılmazdır. Türkiye’de öğrencilerin okullarda uymaları gereken disiplin kuralları, her eğitim basamağına özel olarak hazırlanan yönetmelikle düzenlenmektedir. Ortaöğretim Kurumları Yönetmeliği’nde yer alan disipline ilişkin kurallar, kınama, kısa süreli uzaklaştırma, okul değiştirme ya da örgün eğitim kurumları dışına çıkarma şeklinde cezalar öngörmektedir. Disiplini bozucu eylemin karşılığında hangi cezanın öngörüldüğü, yönetmelikte tek tek sayılmakla birlikte, öngörilemeyen fiiller için de benzer eylemlerde bulunanlara uygun disiplin cezası uygulanacağı yönetmelikte düzenlenmiştir. Öte yandan, çeşitli ulusal veya uluslararası hukuki düzenlemelerle (Anayasa, ilgili Kanunlar, Birleşmiş Milletler Çocuk Haklarına Dair Sözleşme gibi), çocuklara temel hak ve özgürlükler tanınmıştır. 18 yaşını doldurmamış öğrenciler de ilgili hükümler sayesinde temel hak ve özgürlüklere sahiptir. Bunlardan birisi düşünce ve ifade özgürlüğüdür. Ancak söz konusu özgürlük mutlak değildir; çeşitli sınırlar dahilinde kullanılabilir. Fakat çocukların/öğrencilerin okullarda düşünce ve ifade özgürlüğüne sahip olup olmadıkları ve bu özgürlüğün sınırlarının ne olduğu öğretilerde tartışmalıdır. Bu tartışmalar, sadece farklı ülkelerin Anayasa, kanun gibi hukuki metinleriyle değil, çeşitli yargı kararlarıyla da (*Tinker v. Des Moines Independent Community School District* ve *Texas v. Johanson* gibi)

şekillenmektedir. Bununla birlikte Türkiye’de çocukların/öğrencilerin düşünce ve ifade özgürlüğü konusu, toplumsal gelişmeler ışığında ele alınması gereken önemli bir konudur. Türkiye’de çocukların/öğrencilerin düşünce ve ifade özgürlüğüne ilişkin özel düzenlemeler bulunmamakta, söz konusu özgürlüğün kapsamı ve sınırları Anayasa ve ilgili kanun gibi genel düzenlemelerle yapılmaktadır. Öte yandan bahsi geçen özgürlükten çocukların/öğrencilerin hangi sınırlar dahilinde yararlanabileceğini yorumlayan Türk mahkeme kararlarına rastlamak güçtür. Çünkü disiplin düzenleme ve uygulamalarından doğan davalarda çocukların/öğrencilerin sahip oldukları temel hak ve özgürlükler ve sınırları konusunda özgürlükçü değerlendirmeler nadirdir. Bu tür yorumlar için, söz konusu temel hakların ve özgürlüklerin süjesi konumunda olan öğrencilerin konuyla ilgili görüşlerine yer verilmesinin kanun koyucuları, hukuki düzenlemeyi uygulayıcı konumunda olan eğitim yöneticileri ve öğretmenleri ve son olarak uyuşmazlıkları çözen yargıçları olumlu etkileyebileceğine inanılmaktadır. Böylelikle, bu çalışmanın problemini, disiplin düzenleme ve uygulamalarının öğrencilerin düşünce ve ifade özgürlüğüne etkisinin, öğrenci görüşleri doğrultusunda değerlendirilmesi oluşturmaktadır.

Araştırmanın Amacı: Orta öğretim okullarında okuyan öğrencilerin disiplin düzenleme ve uygulamaları bağlamında, düşünce ve ifade özgürlüğünü nasıl değerlendirdiklerini saptamaktır.

Araştırmanın Yöntemi: Araştırma, olgubilim deseniinde tasarlanmış ve yürütülmüş bir nitel araştırmadır. Araştırmanın çalışma grubu, maksimum çeşitlilik örnekleme tekniği doğrultusunda, Ankara ili merkez ilçelerindeki genel kamu ve özel ortaöğretim kurumlarında bulunan tüm öğrencilerden, araştırmaya katılmaya gönüllü 15 öğrenciden oluşmuştur. Çalışma grubunda bulunan 15 öğrenciyle yarı yapılandırılmış görüşme formları doğrultusunda yüz yüze görüşmeler yapılmıştır. Görüşme formları, Milli Eğitim Bakanlığı Ortaöğretim Kurumları Yönetmeliği ile ilgili literatür bir arada ele alınarak oluşturulmuştur. Yazılı metin haline getirilen görüşmeler içerik analizi yöntemi ile analiz edilmiştir. Çözümleme kategorileri (temalar), araştırmanın amacına paralel olarak, ortaöğretim kurumlarında öğrencilerle ilgili disiplin düzenleme ve uygulamalarının, “Düşünce ve İfade Özgürlüğü” boyutunda değerlendirilmesi olarak alınmıştır. Alt kategoriler analizler sırasında tümevarım yöntemiyle belirlenmiştir.

Araştırmanın Bulguları: Öğrencilerin düşünce ve ifade özgürlüğüne ilişkin görüşleri, “söz söyleme” ve “katılım” kategorilerinde toplanmıştır. Öğrencilerin okulda eğitsel amaçlarla, pankart asılması, broşür, bildiri dağıtılması; gazete veya dergi çıkartılması, arma, kol bandı vs. takılması hakkındaki görüşleri, “evet”, “şartlı evet” ve “hayır” kategorilerinde incelenmiştir. Sayılan eylemlerin siyasal amaçlarla yapılmasına ilişkin görüşleri, “evet” ve “hayır” kategorilerinde ele alınmıştır. Son olarak, bu eylemleri yaptığı için öğrenciye kısa süreli uzaklaştırma, tasdikname ile uzaklaştırma veya örgün eğitim dışına çıkarma cezası verilmesine ilişkin öğrencilerin görüşleri, “ceza-yı işlevsiz bulma” ve “hem işlevsiz hem işlevsel bulma” kategorilerinde incelenmiştir.

Araştırmanın Sonuçları ve Öneriler: Araştırmaya katılan öğrencilerin düşünce ve ifade özgürlüğünü kavramsal olarak bildikleri ve çoğunlukla söz söyleme hakkı olarak algıladıkları söylenebilir. Öğrencilerin okulda eğitsel amaçlarla pankart asmalarını, broşür, bildiri dağıtmalarını, gazete veya dergi çıkartmalarını, arma, kol bandı vs. takmalarını, katılımcıların çoğunlukla düşünce ve ifade özgürlüğü kapsamında değerlendirdikleri belirlenmiştir. Ancak sayılan eylemlerin siyasal amaçlarla yapılmasını yaş küçüklüğü, tartışma/çatışmaya yol açması, suç/yasak oluşturması, görüşlerin farklı olması, okulun amaçlarının dışına çıkılması, gruplaşma gerekçeleriyle bahsedilen özgürlük kapsamında değerlendirmedikleri sonucuna ulaşılmıştır. Son olarak, söz konusu eylemleri yapmaları nedeniyle öğrencilerin çeşitli disiplin cezalarını almalarını, katılımcılar çoğunlukla işlevsiz bulmuşlardır. Verilen cezaların ağır olması nedeniyle öğrenciler cezanın amacına ulaşmayacağını ve kınama, uyarı gibi daha hafif cezalar verilmesinin uygun olacağını ifade etmişlerdir. Bu sonuçlar doğrultusunda, disiplin kurallarının öğrencinin yaşına ve olgunluk düzeyine uygun, açık ve anlaşılır olması gerektiği önerilebilir. Disiplin kurallarının öğrencinin temel hak ve özgürlüklerini fiilen kullanılamaz hale getirmemesi gerekir. Düşünce ve ifade özgürlüğüne yönelik sınırlamaların objektif olması önemlidir; aksi takdirde öğrencilerin sahip oldukları hakları kullandıkları için cezalandırılması gibi hakkaniyete uygun olmayan bir sonuç doğabilir. Son olarak söz konusu özgürlüğün etkin biçimde kullanılabilmesi için, düşünce ve ifade özgürlüğüne yönelik sınırlamaların objektif olması ve uluslararası normlara ve demokratik toplum düzenine uygun olması da önem arz eder. Aksi takdirde keyfilik sonucu doğabileceği gibi, toplumsal koşullar nedeniyle temel hak ve özgürlüklerin etkili biçimde kullanılması sağlanamayabilir.

The New Leadership Model of University Management for Innovation and Entrepreneurship

Gamze SART*

Suggested Citation:

Sart, G. (2014). The new leadership model of university management for innovation and entrepreneurship. *Eurasian Journal of Educational Research*. 57, 73-90. <http://dx.doi.org/10.14689/ejer.2014.57.6>

Abstract

Problem Statement: Today's ever-changing educational environment has created a need for new leadership styles that encourage positive change and improvement. In Turkish universities, the most commonly used leadership models are the classic and/or traditional ones, which lead to stagnation in innovation and entrepreneurship. Only a limited number of universities are actively engaged in innovative research activities and achieve success in terms of entrepreneurship and cooperative work with industry. A broad effort is needed to improve cooperation and encourage leadership development.

Purpose of Study: This paper attempts to show and critically analyze the role of leadership models of university management in creating a learning environment for innovation and entrepreneurship.

Methods: Semi-structured interviews were conducted with 42 different faculty members and 12 graduate students over a period of three months at three different universities in Istanbul. Interviews were centered on 6 core research questions. Interpretative Phenomenological Analysis (IPA) was used to analyze the interview outcomes. The data were analyzed using the Atlas.ti 7 software kit.

Findings and Results: A large percentage of the participants, 78%, mentioned that participatory democracy is an important new leadership model that can empower innovation and entrepreneurship. Most respondents, 80%, also identified the important role of intensive collaboration with industry managers. In the transformation of the university, a significant percentage of the participants, 69%, agreed that experts should be frequently consulted and their views taken into

* Dr. İstanbul University, Faculty of Education, E-mail: gamzegazi@gmail.com

consideration. Slightly more than half of the participants, 54%, said that their university could do more to encourage innovation and innovative ideas from students and faculty members; out of the respondents who expressed this idea, the majority, 86%, said that new leadership and management styles would be critical in promoting this change.

Conclusions and Recommendations: As the current study shows, universities that apply new leadership styles create an environment more conducive to fostering entrepreneurship and innovation. Consistent with previous research, the leadership model of university management plays a vital role in universities' readiness to accept innovative and entrepreneurial changes. Universities should adopt new leadership styles instead of using conservative and traditional leadership models that discourage innovation and entrepreneurship. Generally speaking, democratic leadership models are more effective at fostering open innovation. In the new management framework, leaders can create new environments and spaces, such as technology transfer offices, to collaborate with industry.

Keywords: New Leadership Model, University Management, Innovation, Entrepreneurship

Introduction

There has been a problem with regard to leadership and entrepreneurship in Turkish universities. The Scientific and Technological Research Council of Turkey (TÜBİTAK) has only recently decided to rank Turkish universities in terms of entrepreneurship and innovation. The following table provides a clear picture of the current situation.

Table 1

Entrepreneurial and Innovative University Index, 2014 Ranking

Rank	University	Total	Scientific and technological research competence	Intellectual property pool	Cooperation and interaction	Entrepreneurship and innovation culture	Economic contribution and commercialization
1	Orta Doğu Teknik University	83.09	19.6	8.7	22.4	13.8	18.8
2	Sabancı University	81.44	19.5	6.2	25.0	12.5	18.3

Table 1 Continued...

3	Boğaziçi University	76.34	18.5	5.5	24.1	10.0	18.2
4	Ihsan Doğramacı Bilkent University	74.96	19.1	5.2	22.4	12.7	15.6
5	Koç University	73.59	16.0	9.4	24.9	11.3	12.0
...
46	Firat University	29.56	8.7	1.2	4.1	9.8	5.8
47	Kadir Has University	29.19	10.1	0.0	13.9	0.3	4.9
48	Pamukkale University	28.84	7.3	3.5	10.0	2.4	5.6
49	Niğde University	28.25	7.8	2.5	12.3	1.5	4.2
50	Ondokuz Mayıs University	28.22	7.9	2.4	10.8	7.2	0.0

As seen from the table, only a few universities are more or less successful in terms of innovation and entrepreneurship. To make matters worse, only 50 of approximately 200 Turkish universities have been included in this ranking. Therefore, one can easily identify two major problems:

1. Even among the 50 universities that are ranked, there is a huge gap between the first and the last (83.09 versus 28.22);
2. Approximately 150 universities are not even included on the list.

This paper is an attempt to take a critical look at these issues and propose a solution that might offer a significant contribution to the field.

The education given at a university has a direct effect on industry and technology, as these are where educational know-how can be applied to influence the economic development of a country (Ozturk, 2001). This important role of the university is affected positively or negatively by the leadership model of that university's management. For this reason, the leadership model of a university is crucial in promoting innovation and entrepreneurship. Universities must predict the need for change and implement change proactively. Different leadership models play a critical role in guiding such change (Cross, 2012). There are many leadership styles that are increasing in importance not only in universities but also in every area of professional life. As the importance of these styles grows, several questions become

critical for society: "Are the new leadership models suitable for all areas of life or not?" and/or "Are traditional leadership models still the most effective?" However, 'leadership' should first be defined as a concept before considering the question of whether modern leadership styles or more traditional leadership models are better suited for producing university management that promotes innovation and entrepreneurship.

Before analyzing a university management leadership model's effect on innovation and entrepreneurship, it is important to create a perception of leadership. A leader is a person who organizes and holds responsibility over a certain, defined group of people. However, this position is not all that is required to be a leader. A leader must have a plethora of other qualities according to various theories and models. First of all, there is more than one definition of leadership. Different leadership styles incorporate different definitions of leadership, which in turn shape different understandings of managerial roles. There are three managerial roles at senior positions: manager, coach and leader. Thus, leadership is one of these three roles. All senior positions must include coaching in addition to leadership because coaching develops the team and creates the plan (Hays, 2012). Although there are many different definitions of leadership, the Oxford English Dictionary defines leadership as "the dignity, office, or position of a leader, esp. of a political party", "the ability to lead", "the position of a group of people leading or influencing others within a given context", and "the action or influence necessary for the direction or organization of effort in a group undertaking" (leadership, 2014). However, these dictionary definitions are not sufficient to define leadership. Chemers (1997; 2000) defines leadership as a process of social influence in which one person is willing to accept support and help from others to reach a mutually beneficial goal. Classic definitions of leadership, therefore, are reminiscent of "an all-powerful tough-guy of comic book hero proportions" or "a seemingly omniscient individual blessed with unusual foresight" (Smolenyak & Majumdar, 1992, p. 28). These perceptions were valid for hundreds or even thousands of years throughout human history, but today's changing environment changes leadership, as well. Grint (2004, p. 6) indicates four problems with traditional leadership. The first is a 'process' problem: "a lack of agreement on whether leadership is acquired from the personal qualities (i.e., traits) of the leader, or whether a leader induces followership through what s/he does (i.e., a social process)." The second problem involves "position", referring to a dilemma where it is unknown whether the leader is "in charge" or is merely "in front" (i.e., with informal influence). The third problem is 'philosophy'. As Grint describes, this problem is manifested in the question: "Does the leader exert an intentional influence on the behavior of followers, or are their manifest actions determined by context or even attributed retrospectively?" The fourth and final difficulty is 'purity', which is related to whether leadership is embodied in groups or is a purely individual phenomenon.

It can be seen that leadership is a collective notion that touches upon many important social, organizational, and personal processes. These processes include influence and inspiration to work towards group goals, not through coercion but

through personal motivation. Which definition to accept is a matter of choice for the organization and/or the leader, informed by his/her own tendencies, beliefs and the organizational situation but with “an awareness of the underlying assumptions and implications of his/her particular approach” (Bolden, 2004, p. 15). In addition to these definitions of leadership, leadership styles also change from one institution to another. Whereas some institutions require change, others do not need it at all. The needs of an institution affect the leadership role of the senior manager (Hays, 2012). According to Hays (2012, p. 346), there are three important attributes of a good leader. First, a leader must be a good communicator. All staff and all people connected to the organization must be clearly informed of the strategy and its implementation. Hays mentions that “certainty is crucial for confidence, and that a second attribute of a good leader is having a strong awareness of self and other team members, strengths and weaknesses, personality and personal interests. This attribute is essential for optimal team performance achievement.” Hays (2012) also indicates that these leadership characteristics can create a championship team without any individual champions. The third and last attribute of any leader is the ability to delegate decision-making to the team. When the organizational structure is flat, more people contribute to the decision-making process. In other words, the larger the organization, the more important this attribute becomes (Hays, 2012).

To clarify this paper’s topic, new leadership styles need to be considered and implemented. However, just as in the case of leadership, leadership styles also have different definitions. According to Nanjundeswaraswamy and Swamy (2014) “today’s organizations demand effective leaders who understand the complexities of the rapidly changing global environment” (2014, p. 58). If employees have good relationships with their leader and the work that must be done is well structured, the productivity from the leadership is usually high. There are many different leadership styles in the literature. Examples of these leadership models include transformational leadership, transactional leadership, moral leadership, charismatic leadership, adaptive leadership, autocratic leadership, authoritative leadership, laissez-faire leadership, and spiritual leadership (Virkus, 2009). According to Bass and Steidlmeier, there are “two distinct but interrelated ideal types of leadership: transactional and transformational” (1999, p. 187). Transactional leadership involves contingent reinforcement. Bass and Steidlmeier add that in this model, “leaders motivate the followers by their praise, promises, and rewards or correct them by negative feedback, threats, reproof, or disciplinary actions. The leaders react to whether followers carry out what the leaders and followers have ‘transacted’ to do.” Within the model of contingent rewards, “behavior leaders either make assignments or consult with followers about what is to be done in exchange for implicit or explicit rewards and the desired allocation of resources” (Bass & Steidlmeier, 1999, p. 195). Transformational leadership consists of four components. These are “charisma or idealized influence (attributed or behavioral), inspirational motivation, intellectual stimulation, and individualized consideration” (Givens, 2008, p. 20). Subjects “identify with the charismatic leaders’ aspirations” and tend to imitate these leaders in various ways. Leaders’ “charisma or idealized influence is envisioning, if it is transformational” (Mills, 2007, p. 13).

The next step is to analyze the leadership models of universities. Previous research studies have offered two criticisms of university leadership. The first concern is that there has been insufficient development of robust styles of professional management and the second is that conventional thinking and behavior are unchallenged, if not encouraged, because most university leaders have advanced their careers within the system (Scott, 2011). Studies demonstrate that, in particular, transformational leadership produces desirable outcomes for the universities' organizational effectiveness (Pounder, 2001). However, Scott (2011) also mentions that classic university leadership styles can also have advantages; specifically, most senior managers remain firmly embedded in the academic culture. That is, nearly all of the senior managers within this model still regard themselves as academics, albeit inactive ones, as well as managers. This identification gives them a crucial advantage in their internal leadership role. Scott nonetheless suggests that universities may need to develop new leadership models that can provide increased intellectual resources to students and faculty members so that they can better make sense of the complex economic, political, social, and cultural landscape of the twenty-first century world. Highly personalized, informal and widely distributed leadership styles may offer such a model, and a reversion to older models of "collegial" government may offer another. However, neither of these two models is perfect. Whereas the first model may be difficult to adapt to public institutions, the second one is encrusted with memories of irresponsibility and inertia. Nonetheless, either of these models may provide a starting point for the big organization corporate model that is slowly but surely becoming established as the norm in universities (Scott, 2011).

Leadership and entrepreneurship are closely connected in a number of ways. "Entrepreneurship is an expression of efficacy in change being possible whether for profit or not for profit. Especially among recent college graduates and college students, leadership is incrementally being carried out in the form of entrepreneurship" (Ives, 2011, p. 85). Thus, "What is entrepreneurship?" and "What is innovation?" are two questions that must be answered. According to the Merriam-Webster Dictionary, an entrepreneur is "a person who starts a business and is willing to risk loss in order to make money", and innovation is "the introduction of something new" (entrepreneur, 2014; innovation, 2014). In universities today, many students are interested in "being entrepreneurial leaders who affect change in the marketplace and society through creating new for-profit and nonprofit entrepreneurial ventures, instead of serving as more traditional leaders who work to affect change from within existing organizations" (Ives, 2011, p. 87).

In a university whose management operates under a traditional leadership style, improvements in the area of innovation and entrepreneurship will often not be supported. However, research shows that universities that are open to new leadership styles are also more open to innovation. Although there has been intense academic interest in the leadership literature, the studies that examine leadership's relationship with innovation and entrepreneurship in educational settings are limited. The current study aims to extend previous works by analyzing the effect of new leadership styles of university management on creating a learning environment

for innovation and entrepreneurship. Another important and distinguishing factor of this paper is its use of Interpretative Phenomenological Analysis (IPA) as an analysis method to understand participants' perception of the term 'effective leadership models in universities.' With this detailed analysis, the study aims to provide several useful directions for universities and planners of leadership development programs.

In this context, the purpose of this paper is to establish the role of the new leadership and management styles in helping to promote change and encourage innovation and innovative ideas from students and faculty members. Generally speaking, this paper endeavors to answer the following primary questions: "What type of leadership model is needed to encourage innovation?" and "What is required of university management to promote entrepreneurship in universities and increase industry-university collaboration?"

Methods

Research Design:

To collect the opinions of faculty members and graduate students, semi-structured interviews were conducted in three different universities in Istanbul. The interviews were centered on 6 core research questions.

Research Sample

Cluster sampling was utilized in the study. One private and two state universities of Istanbul were selected as major clusters. Within these, further clustering was performed in terms of majors: subjects were randomly selected from three different faculties (Engineering, Medicine and Social Sciences) at these three universities. Faculty members and graduate students are usually involved in innovative research in universities; hence, the sample of the current study consisted of two sub-groups: 42 faculty members and 12 graduate students. The age range was 24-28 for graduate students and 36-55 for faculty members. The distribution of faculty titles with regard to gender and department is given in detail in the following table.

Table 2

of Sample Subgroups with Respect to Faculty and Gender

		Engineering	Social Sciences	Medicine
Faculty members	F	8	9	6
	M	6	8	5
Graduate Students	F	2	2	2
	M	2	2	2

Research Instrument and Procedure

The author conducted a comprehensive literature review while developing the research question. The author was the researcher and the interviewer during the study process, and she generated core questions based on the literature review. Furthermore, five different experts in the field were asked to review the questions to reach a higher level of content validity for the study. Finally, after approval was received from the Ethics Committee, the study was conducted during the spring semester of the 2013/2014 academic year. The researcher visited the selected universities to conduct the interviews. Clear instructions about the procedure and the purpose of the study were provided to the participants. Informed consent, confidentiality and voluntary participation were also ensured. The average interview time was 2 hours. Interviews were conducted face to face, and a voice recorder was used with the permission of the participants.

The 6 core questions were as follows: (1) What type of leadership models do you recommend to increase innovation and entrepreneurship in universities? (2) In your opinion, what is required in the new leadership model of university management? (3) What is required to increase industry-university collaboration? (4) What should be changed first to provide this type of leadership model of university management? (5) What is the role of university management in effective industry-university collaboration? (6) How can new leadership models be applied in universities?

Data Analysis

Atlas.ti 7 software was used for the data analysis. Forty-two faculty members and 12 graduate students were coded as F1, F2, F3... and S1, S2, S3... etc., and descriptive information was obtained. First, the author transcribed all interviews. After the transcription, the written documents were compared with the audio records for further control. Participants' responses were examined in detail and divided into different categories, arranged by content. Interpretative Phenomenological Analysis (IPA) was used to explore how subjects made sense of their experiences and to identify the significance of those experiences to the subjects themselves. Together with IPA, semi-structured interviews are usually used as the best method for collecting data (Chapman and Smith, 2002). In this process, the transcripts were subjected to a case-by-case in-depth analysis. Each transcript was examined repeatedly by taking notes, including interpretations. Afterward, these comments were transferred into themes and connected until coherent thematic integrity was achieved for each participant. After the case-by-case analysis was completed, the themes were connected across participants, and superordinate themes were grouped: (A) "Scientific and technological research competence," (B) "Intellectual property pool," (C) "Cooperation and interaction," (D) "Entrepreneurship and innovation culture," and (E) "Economic contribution and commercialization." Finally, the analysis was translated into a narrative account to provide a reportable format for the paper.

Results

Generally speaking, the results obtained after the aforementioned analysis turned out as expected: a great need for new leadership models was felt and clearly witnessed.

A large percent of the participants, 78% (42 subjects out of 54), mentioned that participatory democracy should be an important component of a new leadership model to empower innovation and entrepreneurship in universities. Regarding the new leadership management, most of the respondents, 80% (43 subjects out of 54), stressed the role of more intensive collaboration with members of industry who play an important role in management. Concerning the transformation of the university, a significant percentage of the participants, 69% (37 subjects out of 54), agreed that the views of experts should be taken into consideration. Nearly half of the participants, 54% (29 subjects out of 54), said that their university could do more to encourage innovation and innovative ideas from faculty members and students. A large majority of the respondents who expressed this idea, 86% (25 subjects out of 29), also stated that new leadership and management styles would be critical in helping to promote this change.

When participants were asked about their recommendations for leadership models that could increase innovation and entrepreneurship in universities, the three keywords most frequently articulated by the participants in different forms were 'participatory democracy', 'participatory leadership' and 'democratic leadership'. Participants mentioned the importance of respect and engagement in university management to achieving real development and growth. An absolute majority of participants agreed that the universities at which they study or work at should possess more democratic and more effective models of leadership. Because educational leadership has not yet been sufficiently conceptualized by the academic community, the variance for the codes derived during interviews from the participants' answers to the first question was high. However, there was a remarkably high consistency in the statements in terms of democratic practices to be applied by the university management. Participants expressed their belief that the most effective and productive learning environments could be provided via better internal and external contributions. They wanted to take a more participatory role in the decision making process and also sought to involve other people in the process. For those participants who did not directly use the word 'participatory democracy' in the university management leadership model, 22% (12 out of 54 participants), there was again a significant emphasis on the influence of others on the leader's decision. However, there was a hierarchy between answers in terms of how exactly how participatory the leadership style should be. There was only 1 participant who supported the view that leaders should make autocratic decisions. Four participants said that a leader should listen to feedback and talk with others but should ultimately make the final decision by him- or herself. Four participants favored a model in which a team processes decisions, yet the final decision still belongs to the leader. Three participants mentioned a joint decision-making style carried out by a

team. This range of responses indicates that although some participants did not use the most common keywords at all, they nonetheless were thinking of the same or closely related characteristics of participatory democracy in the university management leadership model.

In the second question, participants were asked their opinion concerning the requirements for a new leadership model of university management. The most frequently appearing keyword was 'intensive collaboration'. Both faculty members and graduate students mentioned the shared responsibility of the university management in action for faculty members, students, managers, other stakeholders, government, and even the public. A large percentage of the participants mentioned the link between academia and industry in promoting innovation and entrepreneurship. One of the participants insistently emphasized the advantages of this type of collaboration, saying, "University-industry collaboration is a must in the new era. It enhances the university's research mission and advances its position as a leading source of innovation. It also provides a public service, benefits the economy by attracting investment, creates jobs and enhances the quality of life. Academicians who do not see this will always be behind the times."

In the third question, participants were asked, "What is required to increase university and industry collaboration?" Participants mentioned the role of industry managers in university management leadership. Some participants gave examples from Turkey, where many company owners have built private (foundation) universities, thereby empowering intense collaboration. Faculty members from private universities mentioned corporate and institutional partnerships. One participant said, "This is an excellent opportunity for the industry to have an innovative partner, to access good research, different pools of knowledge, talented future employees and to save costs on Research and Development." In general, participants held the view that companies must actively embrace universities. In their opinion, the differences between industry and academia can be transformed into a significant advantage through collaboration.

"What should be changed first to provide this type of leadership model of university management?" was the fourth question. As can be inferred from the narratives, the communication between industry and academic teams is at a lower level than expected. The process is generally hampered and slowed by bureaucracy. For this question, the keywords derived from a large percentage of the participants' answers were 'communication' and 'sharing ideas'. Subjects stress the importance of face-to-face encounters in which both university management and stakeholders can enthusiastically pass along their ideas. 'Networking' is yet another keyword that is frequently found in the answers to the fourth question. Between universities and the business community, networking is a critical component of an efficient innovation. Some of the participants gave several national and international examples of established networking tools that create links between universities, business and research technology organizations. They thought that these mechanisms should be reviewed and updated as innovations that can change communication capabilities and expectations. According to the data analysis, the participants frequently articulated responses within many subcategories of the superordinate theme

‘communication’. Integrated as meaningful units, they appear as the following: ‘common understanding of the objectives’, ‘constant contact in both directions’, and ‘sharing operational and strategic decisions’. Participants also evaluated open communication as a part of problem solving. One faculty member gave an example: “Universities cannot deliver all the services that industry desires. It is natural because the resources are scarce and sometimes the expectations do not match the reality. It can only be solved through clear and open communication between the principals at an executive level. It is essential that this be done with the mutual respect of the needs and constraints of the other party. In order to achieve this for industry–university collaboration, transparency of the responsibility at an executive level is a prior necessity.”

Participants were asked about the role of university management in effective industry–university collaboration. The analysis of this question revealed that universities should adopt collaboration and transform as quickly as possible. To clarify this point, the industrial transformation of universities aims to encourage a larger number of collaborative Research and Development projects that can address challenging issues faced by industry today through innovative research. The most frequent keywords related to the expected goals from transforming university management presented a coherent unity for this question. Final statements obtained from the data analysis were ‘developing an innovative mindset for students and faculty members’ and ‘adopting more collaborative and cross-disciplinary problem solving methods’. Faculty members said that they were not always in full agreement with regards to innovative collaboration decisions. This problem appears to be one of the most prominent of those hindering the transformation process. The majority of participants, 87%, agreed that the focus of new leadership models of university management should be on promoting this change.

Lastly, participants evaluated how new leadership models can be applied in universities. The most frequently appearing keyword was ‘experts’. According to the narratives derived from the interviews, expertise in the field is required to lead this transformation process. In terms of ongoing collaboration, universities generally operate within a framework of multiple partnerships that is hard to manage. The need for experts simultaneously implies the engagement of university leaders in the transformation. Fostering leadership performance and exploring how to build leadership and management skills and effect positive institutional change are essential in terms of widening the networks of leaders in the business world, academia, and other segments of society. At this point, innovation depends on the ability of university and industry experts to work together across a number of disciplines. Due to the flexible frame of the semi-structured interview method, this question also reveals the views of faculty members on another topic, i.e., the promotion of a multidisciplinary approach to research and learning to pave the way for partnerships between industry researchers and academicians based on multidisciplinary research. Through these partnerships, universities and industry can pursue solutions to complex and systemic problems that require cross-disciplinary expertise.

Discussion and Conclusions

The goal of this paper was to discuss whether university management who use new leadership styles are more open to innovation and entrepreneurship. Findings confirmed the view that new leadership styles are more conducive for making improvements. New leadership models promote positive change among people within the university and even beyond the institutional level.

However, as signified by Simsek (2013), this change as well as new theories and concepts may not be easy to apply since new liberal ideas, even though very instrumental in promoting change in industry, have so far had a relatively weak influence on education system. On the other hand, a profound and somewhat revolutionary change of university leadership has been needed, and the importance of promoting this change has been growing.

Transformational leadership also has a great impact on the higher education itself in terms of learning and teaching (Quinlani, 2014). This aspect usually skips the attention of the researchers even though it is of vital significance since the transformational leadership is directly linked to a holistic student development and naturally leads to closer ties with markets. Even though the major constituents of such leadership are arguable, mainstream trend in this area of educational research shows that transformational leadership in its essence is very close to the idea of comprehensive collaboration of industry and universities.

According to many pieces of academic literature on leadership, a key aspect of many transformational leaders is that they are oriented towards the future, not the past. However, striving to determine and implement the desired future direction of an organization is essentially an ambiguous and uncertain undertaking (Simsek, 2013). Therefore, it is also fundamentally a risky venture, and as a result, transformational leaders are the most effective and work best in more chaotic environments. This is contrasted with transactional leaders, who best function in calm and stable periods of an organization where there are clearly defined and commonly agreed-on standards and rules. However, many would agree that the field of higher education is not in such condition now.

As was shown before, since most of the university staff members have reached top positions within the boundaries of the present academic environment (Scott, 2011), challenging this conventional framework becomes difficult. It may demand serious reconsideration of the basic principles and rules established for the area long ago. Active involvement of the experienced academicians, despite the fact that they may tend to stick to the old framework, is absolutely necessary, as many of them are experienced leaders with valuable insight to give. This was also indirectly confirmed by the participants of this study, as they mentioned listening to experts as one of the major factors necessary to promote change.

Taking the aforementioned arguments into consideration, consistent with the literature review, the current study results show several critical points that cannot be overlooked. They can be underlined as follows:

First, as a major component for empowering innovation and entrepreneurship within higher education, participatory democracy was on the top of the list for creating a successful future leadership model. Achieving a democratic environment at an organizational level may positively affect leadership development by expanding awareness of the primary goals and contributions expected of the organization.

Second, new models of leadership management have no future without a thoroughgoing collaboration with industry management members. With respect to this finding, as a supportive argument, a strong need for qualified experts in the area was emphasized by the participants of the study; thus, expert opinion should play a major role in the process of the higher education transformation. This may have far-reaching consequences for the higher education at large, leading to re-shaping or even forming absolutely new entities and branches of management mentorship within universities. Possible future prospects of such innovations have already inspired the creation of annual Entrepreneurial University Leadership Programs and similar projects (Gibb, Haskins, and Robertson, 2013).

Third, positive and constructive changes in terms of promoting innovation within higher education cannot be achieved without a significant increase in motivation. Participants said that universities should encourage innovation and innovative ideas from students and faculty members. They emphasized the role of new leadership and management styles in helping to promote this change. The same values and goals may have a better chance of becoming embedded when they are signaled implicitly from the top, which can motivate the underlying layers of management to react in a creative way. For the universities, executive officers and senior teams can assume an important role by promoting organizational culture in a positive way in terms of innovation and entrepreneurship at their universities.

Some of the narratives have implied that there is a rather unusually high demand not only for new leadership models but new leaders as well. Thus, not only has this paper showed that universities need new leadership to strengthen their ties with real sectors of economy, but it has also indirectly proven that the demand for sustainable entrepreneurs, people who have the capacity to facilitate development, has also been growing (Lans, Blok, and Wesselink, 2014). Such disciplines like entrepreneurship education and education for sustainability must be integrated in order to achieve a stable and constantly developing framework for sustainable entrepreneurship. Such approach can be very instrumental in fostering entrepreneurial spirit in the higher education.

More generally speaking, narratives from the research participants expressed how they perceived effective leadership in universities and also emphasized the important role they attributed to intensive collaboration with members of industry, particularly those with important management positions. It is important to gather relevant information that affects strategy and operations. Leaders should listen to staff and external stakeholders. Finally, all participants of the study agreed that it

might be quite a struggle to remain relevant and continuously improve and move in a positive direction towards innovation in higher education institutions.

One of the areas that can be suggested with regard to the continuation of the present research is the gender issue within the leadership. As Michael states (2013), women tend to exhibit a predominantly transformational type of leadership in higher education. They usually take other paths and develop highly flexible forms of leadership that can be worth of studying as a separate phenomenon or in connection with the issues discussed in the current study.

As a limitation of this study, its relatively small sample of faculty members and students can be mentioned. Therefore, it is recommended that future studies be conducted that target larger samples from different regions of Turkey, in order to draw a better picture of leadership issues in other parts of the country. Gathering data from a larger research sample and gaining insight on particular challenges that leaders face within the university environment can provide further guidance for leadership development in universities.

References

- Bass, B. M. & Steidlmeier, P. (1999). Ethics, character, and authentic transformational leadership behaviour. *Leadership Quarterly*, 10 (2), 181-217.
- Bolden, R. (2004). *What is leadership?* University of Exeter.
- Chapman, E. & Smith, JA. (2002). Interpretative phenomenological analysis and the new genetics. *Journal of Health Psychology*, 7, 125-130.
- Chemers, M. M. (1997). *An integrative theory of leadership*. Lawrence Erlbaum Associates.
- Chemers, M. M. (2000). Leadership research and theory: A functional integration. *Group Dynamics: Theory, Research, and Practice*, 4 (1), 27-43.
- Cross, S. E. (2012). A leadership model for the research university. *3rd International Conference on Leadership, Technology and Innovation Management*. Atlanta, GA: Elsevier.
- Entrepreneur, n. (2014). *Merriam-Webster dictionary*. Retrieved May 31, 2014, from Merriam-Webster Dictionary:
<http://www.merriam-webster.com/dictionary/entrepreneur>
- Gibb, A., Haskins, G., & Robertson, I. (2013). Leading the entrepreneurial university: Meeting the entrepreneurial development needs of higher education institutions. In *Universities in Change* (pp. 9-45). Springer New York.
- Givens, R. J. (2008). Transformational leadership: The impact on organizational and personal outcomes. *Emerging Leadership Journeys*, 1, 4-24.
- Grint, K. (2004). *What is leadership? From hydra to hybrid* [Working paper]. Saïd Business School and Templeton College, Oxford University.

- Hays, R. (2012). Leadership. *Blackwell Publishing Ltd*, 345–347.
- Innovation, n. (2014). *Merriam-Webster dictionary*. Retrieved May 31, 2014, from Merriam-Webster Dictionary: from <http://www.merriam-webster.com/dictionary/innovation>
- Ives, A. (2011). Entrepreneurship education as a new model for leadership education. *Journal of Leadership Studies*, 5 (2), 85-88.
- Lans, T., Blok, V., & Wesselink, R. (2014). Learning apart and together: towards an integrated competence framework for sustainable entrepreneurship in higher education. *Journal of Cleaner Production*, 62, 37-47.
- Leadership, n. (2014, March). *OED online*. Retrieved May 29, 2014, from Oxford University Press: <http://www.oed.com/view/Entry/106604?redirectedFrom=leadership&>
- Lewis-Beck, M. S., Bryman, A. E. & Liao, T. F. F. (2003). *The sage encyclopedia of social science research methods*. Sage Publications.
- Michael, R. (2013). *Women in leadership in higher education: Leadership styles and the* (Doctoral dissertation, TREVECCA NAZARENE UNIVERSITY).
- Mills, G. E. (2007). *Transformational leadership and employee retention: An exploratory investigation of the four characteristics*. (Unpublished Degree of Doctor in Philosophy). Capella University, U.S.
- Nanjundeswaraswamy, T. S. & Swamy, D. R. (2014). Leadership styles. *Advances In Management*, 7 (2), 57-62.
- Ozturk, I. (2001). The role of education in economic development: A theoretical perspective. *Journal of Rural Development and Administration*, 33, 39-47.
- Pounder, J. S. (2001). “New leadership” and university organisational effectiveness: Exploring the relationship. *Leadership & Organization Development Journal*, 22, 281 – 290.
- Quinlan, K. M. (2014). Leadership of teaching for student learning in higher education: what is needed?. *Higher Education Research & Development*, 33(1), 32-45.
- Scott, P. (2011). Leadership in universities. *The International Journal of Leadership in Public Services*, 7 (3), 229-234.
- Simsek, H. S. (2013). Transformational leadership in educational context: a fantasy of education scholars (Selected Keynotes). *Eurasian Journal of Educational Research*, 51, 1-6.
- Smolenyak, M. & Majumdar, A. (1992). What is leadership? *The Journal for Quality and Participation*, 15 (4), 28.

Stehman, S. V. (1997). Estimating standard errors of accuracy assessment statistics under cluster sampling. *Remote Sensing of Environment*, 60 (3), 258-269.

Virkus, S. (2009). *Leadership models/approaches*. Retrieved from Leadership Models:

http://www.tlu.ee/~sirvir/Leadership/Leadership percent200M Models/leadership_modelsapproaches.html

Üniversite İnovasyon ve Girişimcilik Yönetimi Yeni Liderlik Modeli

Atf:

Sart, G. (2014). The new leadership model of university management for innovation and entrepreneurship. *Eurasian Journal of Educational Research*. 57, 73-90.

<http://dx.doi.org/10.14689/ejer.2014.57.6>

Özet

Problem Durumu: Günümüzün sürekli değişen koşulları, liderlik kavramını da değiştirmiş ve yeni liderlik modellerinin adapte edilmesi ihtiyacını doğurmuştur. Teknolojideki gelişmeler eğitim öğretim alanında olumlu değişim ve gelişmeler sağlamış fakat inovasyon ve girişimcilik konusunda açık olan üniversiteler bundan faydalanırken, Türkiye'nin genelinde görülen geleneksel liderlik modellerinin sürdürüldüğü okullar çağa ayak uydurmaktan uzak kalmışlardır.

Araştırmanın Amacı: Bu makalede, üniversite yönetimi liderlik modelinin ileri teknoloji, yenilikçi ürün ve hizmetlere artan talepten etkilenen inovasyon ve girişimcilik ortamlarının yaratılmasındaki etkisini tartışmak amaçlanmıştır.

Araştırmanın Yöntemi: Bu çalışma, yenilik ve girişimciliğin önemli ölçüde geliştiği üniversitelerdeki liderlik modellerini eleştirel bir şekilde sorgulamaktadır. çalışmanın amacına yönelik olarak, üniversite yenilik ve girişimcilik yönetimi yeni liderlik modeli, oldukça rekabetçi bu küresel ekonomilerdeki yeni rollerini iyice anlamak için açık bir şekilde araştırılmıştır. Çalışmada, üç aylık bir süre boyunca 42 farklı öğretim üyesi ve 12 yüksek lisans öğrencisi yarı yapıllı bir şekilde görüşmeler düzenlenmiştir. Katılımcılar 19 erkek 23 kadın öğretim üyesi; 6 erkek 6 kadın yüksek lisans öğrencisi olmak üzere 3 farklı departmandan gelmektedir (Tıp Fakültesi, Mühendislik Fakültesi ve Sosyal Bilimler Fakültesi). Öğretim üyelerinin büyük çoğunluğunu profesörler oluşturmaktadır. Öğrenciler için yaş aralığı 24-28 iken öğretim üyeleri için 36-55 arasında değişiklik göstermektedir. Bu yarı yapıllı görüşmeler, 6 ana araştırma sorusu etrafında şekillenmiştir. Sorular şu şekildedir: (1) Üniversitelerde girişimcilik ve inovasyonun artması için ne tür liderlik modelleri önerirsiniz? (2) Size göre yeni Liderlik yönetiminde neler gereklidir? (3) Özellikle endüstri-üniversite işbirliğinin artması için neler gereklidir? (4) Üniversite yönetiminde böyle bir liderlik modelini adapte edebilmek için en başta nelerin değişmesi gerekmektedir? (5) Endüstri-üniversite işbirliğinin etkin olduğu bir yönetimde üniversiteye düşen rol nedir? (6)

Bu yeni liderlik modelleri üniversiyelerde nasıl uygulanabilir? Yüzyüze yapılan bu görüşmelerde katılımcıların izni alınarak ses kayıt cihazı kullanılmıştır. Yorumu dayalı fenomenolojik analiz ve Atlas.ti 7 ile toplanan veriler analiz edilmiştir.

Çalışmanın analizi sırasında öncelikli olarak tanımlayıcı istatistik verileri elde edildi. Sonrasında tüm görüşmelerin tek tek yazar tarafından deşifresi yazıldı. Yazılı materyaller ses kayıtları ile karşılaştırılarak kontrolleri yapıldı. Her bir katılımcının cevapları vaka bazlı olarak detaylı şekilde incelendi ve içeriğe bağlı farklı kategoriler oluşturuldu. Bir vaka analizi sırasında alınan notlar ve yorumlar değerlendirilerek tek bir katılımcı için tematik bütünlüğe ulaşılan kadar üzerinde tekrar tekrar çalışıldı. Vaka vaka yapılan analizler tamamlandıktan sonra katılımcılar karşılaştırılarak üst kategoriler oluşturuldu ve bir tabloya yerleştirildi. Son olarak da analiz tablosu rapor edilebilir bir format sağlamak amacıyla öyküleyici formata çevirildi.

Araştırmanın Bulguları: Katılımcıların çoğunluğu (percent 78), yenilik ve girişimciliği güçlendirmek için katılımcı demokrasinin yeni liderlik modeli kadar önemli olduğuna değinmiştir. Yeni liderlik yönetiminde çoğunluk (percent 79), yönetimde büyük rol oynayan endüstri üyeleriyle yoğun işbirliğine açıklık getirmiştir. Katılımcıların önemli bir oranı (percent 69), üniversitelerin dönüşümünde uzmanların dikkate alınması gerektiği konusunda hemfikirdir. Katılımcıların neredeyse yarısına (percent 54) göre üniversiteler, öğrencilerden ve öğretim üyelerinden gelen yeniliği ve yenilikçi düşünceleri daha fazla teşvik edebilir. Bu düşüncüyü dile getirenlerin çoğunluğu (percent 87), yeni liderlik ve yönetim tiplerinin, bu değişimi desteklemeye yardımcı olma konusunda son derece önemli olabileceğini söylemiştir.

Katılımcılara liderlik modelleri ile ilgili önerileri sorulduğunda farklı formlarda en sık kullanılan anahtar kelimeler şunlar oldu: 'katılımcı demokrasi', 'katılımcı liderlik' veya 'demokratik liderlik'. İkinci soruda katılımcılara yeni liderlik modeli için nelerin gerekli olduğu sorulduğunda 'yoğun işbirliği' en sık kullanılan anahtar kelimeler oldu. Katılımcılar sıklıkla endüstri-üniversite işbirliğinin çağa ayak uydurma açısından önemini vurgularken pek çoğu bu işbirliğinin üniversite, endüstri ve ekonomi açısından avantajlarını dile getirdi. Bu yeni liderlik modelinin sağlanması için değiştirilmesi gereken ilk şey, katılımcılara göre iletişimin artırılıp fikirlerin daha çok paylaşılması şeklinde oldu. Üniversitelerin içte ve dışta iletişim ağını kuvvetlendirmesi ve düşüncelerin açıklıkla ifade edilmesi hem inovasyon ve girişimciliği desteklemek açısından hem de yönetimle ilgili problem çözümündeki etkisi açısından sıklıkla dile getirildi. Endüstri-üniversite işbirliğinde üniversiteye düşen rol sorulduğunda katılımcıların büyük bir yüzdesi üniversitenin en kısa sürede buna adapte olup dönüşümünü tamamlaması gerektiğini vurguladı. Son olarak bu liderlik modellerinin nasıl uygulanabileceği konusunda en sık kullanılan anahtar kelimelere göre katılımcılar bunun konusunda uzman kişiler vasıtasıyla yapılabileceğini düşünmektedir.

Araştırmanın Sonuçları ve Önerileri: Bu çalışmanın amacı, yeni liderlik türlerini kullanan üniversite yönetimlerinin inovasyon ve girişimciliğe daha açık olup

olmadıklarını tartışmak olmuştur. Elde edilen bulgulara göre yeni liderlik modellerini uygulayan üniversiteler inovasyona daha uygundur. Geleneksel liderlik modelinin sürdüğü üniversitelerde ise inovasyon ve girişimcilik genellikle desteklenmemektedir. Liderlik konusunda literatürde pek çok çalışma yer alsa da yeni liderlik modellerinin eğitim alanındaki rolünü inceleyen çalışmalar kısıtlıdır. Bu çalışmada ise adapte edilen liderlik modelinin üniversitelerde eğitim, öğretim ve yönetim alanlarındaki etkileri derinlemesine incelenmiştir. Makalenin ayırt edici bir diğer özelliği ise yöntem olarak yoruma dayalı fenomenolojik analizi kullanarak katılımcıların üniversitelerde etkin liderliğin gereklilikleri konusundaki algılarını onların bakış açısından analiz etmesidir.

Literatür ile tutarlı olarak çalışma bulguları üniversite yönetimi liderlik modellerinde katılımcı demokrasinin önemini ortaya çıkarmıştır. Bu demokratik ortamın sağlanması durumunda liderlik gelişimi daha bütünsel bir noktaya ulaşabilir ve üniversiteki etkileyen tüm faktörlerin farkındalığının artmasına yardımcı olabilir. Bir diğer belirgin bulgu ise katılımcıların inovasyon ve girişimcilik alanında motivasyon ve desteğin gerekliliğini vurgulaması şeklindedir. Bu konuda üniversite yönetimine önemli görevler düşmektedir. Organizasyonel kültürü olumlu yönde oluşturma bu anlamda atılacak ilk adım olabilir. Katılımcıların dile getirdiği endüstri üyeleriyle işbirliği konusunun ötesinde liderler üniversitenin içinde ve dışında işbirliği içinde olup stratejik ve işlevsel kararlarda diğerlerinin fikirlerini dinlemelidir. Böylece hem liderler bireysel anlamda hem de üniversiteler organizasyonel anlamda sürekli irtibat halinde kalabilir ve gelişimlerini sürdürebilirler.

Anahtar Sözcükler: Yeni liderlik modeli, üniversite yönetimi, inovasyon, girişimcilik

Investigation of Vocational Interest and Preference in Terms of Gender and Socio-economic Status

Kaan Zülfikar DENİZ*

Ersin TÜRE**

Aslı UYSAL***

Tuba AKAR****

Suggested Citation:

Deniz, K. Z., Türe, E., Uysal, A., & Akar, T. (2014). Investigation of vocational interest and vocational preference in terms of gender and socio-economic status. *Eurasian Journal of Educational Research*, 57, 91-112
<http://dx.doi.org/10.14689/ejer.2014.57.1>

Abstract

Problem Statement: Individuals tend to prefer a vocation in order to reach their targets such as leading a life, nutrition, housing, being safe, having a good position in society etc. It is a task of the adolescence period to choose a vocation which is for some the most important step of the life, while for some others it is a rather important step in life however that can be changed and made flexible any time. All studies which are examined, show that vocational interest and vocational preference are affected by a great deal of variables.

Purpose of Study: In this study, the relations were investigated between the gender, socio-economic status (SES) variables and the vocational interest and the vocational preference.

Methods: Within this study, the long (OFII) and short form (OFII-SF) of the Occupational Field Interest Inventory were used. In total 812 students, who were studying at the preparatory classes of Ankara University (N=495) and Yıldırım Beyazıt University (N=317) during the 2011/12 spring semester, took part in this study. Before the analyses were made, the raw scores obtained via OFII and OFII-SF were converted into z score.

*Corresponding author : Assist. Prof. Dr., Ankara University, Institute of Educational Sciences, Ankara, Turkey. zlfkrdnz@yahoo.com

**Research Assist., Ankara University, Faculty of Educational Sciences, Ankara, Turkey. ersinture@gmail.com

***Curriculum Specialist, B.Ü. Ayşeabla Private Schools, Ankara, Turkey. asli.uysl@gmail.com

**** Curriculum Specialist, Ankara University, Turkey. tubakoglu@gmail.com

For the analysis of the data chi-square, independent samples t test and one way ANOVA statistic techniques were used.

Findings and Results: At the end of the study, it was seen that there is significant difference with regard to the gender of the individuals in terms of the fields which they study and their interests. According to the findings that were obtained as per SES, it was found out that the departments of current higher-education and vocational interest (except for mathematics) of the individuals do not differ significantly.

Conclusions and Recommendations: It was found significant that the point differences with regard to OFII sub-dimensions are in favor of women for Psychology, Health, Law, Foreign Languages and Visual Art; in favor of men for Computer, Political and Financial Sciences, Engineering, Agriculture-outdoor and Physical Science. It can be asserted that these findings are in conformity with the gender-vocational interest studies. Many studies show that the interest/preference of the women for social vocations, which require working with people, is higher, while interest of men for the vocations which require working with objects and abstract concepts is higher. It was recommended, while measuring vocational interests that the gender variable is to be considered and new studies on SES variable carried out.

Keywords: Vocational interest, Gender, SES, Occupational Field Interest Inventory (OFII).

Introduction

Individuals tend to prefer a vocation in order to reach their targets such as leading a life, nutrition, housing, being safe, having a good position in society etc. It is a task of the adolescence period to choose a vocation which is for some the most important step of the life, while for some others it is a rather important step in life however that can be changed and made flexible any time. Choosing a vocation means choosing a life style.

Vocational preference is described as the process of choosing one vocation among a group of vocations. While making the vocational preference, the individual spends a time in which he/she determines suitable/unsuitable vocations for himself/herself. There are subjective and objective factors in vocational preference. Subjective factors are interest, demand, skill and social status; objective factors are payment, income and promotion, it means social security and economic return (London, Würzburg & Berne 1972). The three variables that become prominent in the vocational preference of the individual are skill, value and interest. While making a choice, the individual compares his very own details (interest, life style, values, skill etc.) with the options at hand (vocations, sectors, location of companies, social security, etc.) and by means of this matching he/she chooses a vocation (Varçın et al. 2005).

Kuzgun (2000) describes that the preference of a vocation is arraying the vocations in terms of desirability and emphasizes that factors like interest, social reputability and income effect this preference. Gottfredson's theory (1981) is one of

the more applied theories in the field of vocational goal construction. This theory is presented and discussed in terms of its fundamental concepts: Self - ego, vocational perception, cognitive map of preference, vocational accessibility and compromise process, vocational choice and vocational aspiration (As cited in Külahoğlu, 2001). It is emphasized in International Association for Educational and Occupational Guidance (2002) that in determining the aims of the individual, the values, interests and career decision should always be reviewed.

Strong (1943), who has carried out the very first study about the interest, describes interest as follows: "The 'liking', 'not liking' or 'being indifferent' reaction of a person against another person, thing or action" (As cited in Herr & Cramer, 1996). Interest, according to Psychological and Psychoanalytic Terms Dictionary (A Comprehensive Dictionary of Psychological and Psychoanalytic Terms), is described as distinguishing an object or situation, or the attitude or sense in form of self-orientation (As cited in Savickas, 1999). There are many studies and scales developed about interest too. (Balkıs, 2004; Barak & Cohen, 2002; Gasser, Larson, & Borgen, 2007; Helledey, Zytowski, & Fouada, 2004; Rottinghaus, Coon, Gaffey, & Zytowski, 2007). One of them is vocational interest.

Vocational interest is regarded as an aspect of interest concept. Holland (1985) describes the vocational interest as a part of personality and indicates that it is the sign of the personality at work, lectures at school, hobbies, leisure time activities and preferences. Many psychometrists have described vocational interests as "I like", "I do not like" or "Does not matter" reactions given by the individual in situations concerning vocations and facilities (Savickas, 1999). Despite different points of view, there are many common points in describing vocational interests. Based on this information, interest may be defined as internal process when upon request of the individual he/she pays attention to an object without extra effort, keeps it for long, is aware of it and converts this into reaction and behavior.

Low and Rounds (2007) indicate that the vocational interests are affected by characteristics like identity formation; cognitive, social and emotional evolution, hormonal, psychological and physical changes so much as it is affected by family, school and peers during the adolescence. Exclusive to Turkey, individuals make their choices according to the point they get from the university entrance exam, that is to say in terms of information and skill rather than interest.

Vocational interests can be determined in two different ways as disclosed and measured ones. The disclosed information are in the form expression of how the individual is directly interested in which job. The measured interests are put through via inventory. Although there are different methods used in measuring vocational interests, the most common method is measuring via interest inventory (Silvia, 2006), Strong Interest Inventory-SII, Kuder Preference Record-Occupational, Kuder Career Search - KCSonline, Holland's (Self-Directed Search-SDS), Kuzgun's Academic Self-Concept Scale (ABKÖ), Kuzgun's Self-Assessment Inventory (KDE), Newspaper News Test (GHT) and Occupational Field Interest Inventory (OFII) are some of the scales used in determining vocational interests. Their usage at the present time on computers helps their proliferation. In terms of gender (Betz & Schifano, 2000; Eccles, 1993), racial/ethnic group (Leong, 1995) and socio-economic status (SES) (Bandura et

al., 2001) there are differences in vocational interest types (As cited in Low, Yoon, Roberts, & Rounds, 2005).

Gender is the most explicit variable in distinguishing interests (Erdil, 2006; Lippa, 1998, 2005; Low et al., 2005; Rounds, 1995; Sayın, 2000; Su, Rounds, & Armstrong, 2009; Tay, Drasgow, Rounds, & Williams, 2009). Many studies, literature reviews and meta analyses show that men are interested in vocations concerning objects, whereas women are interested in vocations in which you work with people. In the same manner, their choices are parallel to their interests (Lippa, 1998, 2005; Low et al., 2005; Rounds, 1995; Su et al., 2009; Tay et al., 2009). Low and Rounds (2007) state that interests are in line with gender. According to these studies, women and men tend to be interested in activities and vocations which are stereotyped for their gender. Low et al. (2005) have reached the result that there are differences between women and men for changeability of interests. In their study, they came through that the interests of men are more stable than women's.

On the other hand, Anderson, Tracey and Rounds (1997) compared women and men in the sense of General Occupational Field which is included in Strong's Interest Inventory and observed no significant difference between them. Another study, though, is in contrast with this result. According to Rounds and Day (1999) women and men are put in for various fields in the business life in terms of tradition, social structure and opportunities.

Deng, Armstrong and Rounds (2007) came up with the conclusion that gender has determinative affect on vocational interest structures. These researchers analyzed the gender difference in respect of Holland's six-personality types and together with the results they recommended realistic vocations for men and social ones for women. In view of Holland's theory, people with realistic personality are interested in object-oriented vocations.

The most evident vocation in this dimension is engineering (Choices Interest Profiler and UCF Majors, 2012). Similar to this result, according to Armstrong and Rounds (2008), cited Reardon, Vernick and Reed (2004) as finding the vocations in labor market where women and men are dominant disperse unequally in Holland's six-personality model. In labor market, men tend to be dominant in investigative and realistic ones, whereas women tend to be dominant in social and conventional vocations. As a result of the 35-year longitudinal study carried out by Lubinski and Benbow (2006); although women with high levels of math proficiency prefer mathematics and science fields, they cannot build a career as successful as men in these fields. Nevertheless, they observed that women build career in vocations like health, law and social sciences which require multidimensional (math ability as well) abilities. Among the individuals who were monitored, men and women are equally happy in their preferences in mid-thirties and successful in direction of aim determined by themselves.

SES is another variable in determining vocational interests. Lent (2005) states that the career of the individual is affected both by the role in her/his life and with SES, and according to the information gained from Lent, the career life of the individual is built in the way of her/his past life. Lent states that SES is another important variable in vocational interests together with ethnic origin, genetics and physical health.

According to Patton and McMahon (2006) cited Holland (1998) as finding, gender and SES are important variables for vocational interest and preference. Bütün (1998) reached the conclusion in his/her study that SES is also effective in determining the interests of female students together with the vocation and educational background of mother and father and birth order. Erdil (2006), on the other hand, came through in his/her study on females and males that the number of the attended university entrance exams, educational level of mother-father, income level and type of the graduated high school are effective on vocational preference. Strong (1943) states that interests are permanent and not affected sufficiently by the vocational education and experience for estimating future behaviors. Yet, McArthur (1954) assumes that this relation seems to be changeable by SES (Betsworth & Fouad, 1997).

All these studies show that vocational preference and interest are affected by a great deal of variables. One's professional interests should affect their career choices. However, a wealth of research observations made in Turkey indicated that individuals most often choose their occupation based on their idiosyncratic skills, as well as, from influences in their personal environment. In this study, the relations were analyzed between the gender, socio-economic status (SES) variables and the vocational interest and the vocational preference. To indicate vocational interest "OFII" was used. In this investigation, a determination was made regarding participants' vocational choices based on their observed majors of study or their job profession. As a result, in this investigation the concept of professional interest was measured using OFII in 14 dimensions, and the occupational preferences were limited to the participants' university department.

Method

Research Design

This study focuses on examining the relations between the variables, therefore, it can be classified as a correlational survey model. Correlational survey model, which states the relations gained by comparison and correlations between variables, is one of the descriptive survey models (Karasar, 2011). In correlational survey model while determining the survey problem and the variables, it is necessary to get support from previous survey results, the theory, expert views or experiences in order to determine whether the relationship is worth examining or not.

Participants

In total 812 students, who study at preparatory classes, took part in this study. 495 (61.0 %) of the students are from Ankara University and 317 (39.0 %) of them are from Yıldırım Beyazıt University. Of the participants, 431 (51.1%) were female, 380 (46.9%) were male and one person did not will to state gender. When observed with regards to ages, 702 (86.6 %) of the participants were at the age of 19-20, 90 (11.1 %) of them were above 20 and 19 (2.3 %) of them were 18 and below. One participant did not will to state age.

Table 1 shows the distribution of the participants in respect of their departments of current higher-education.

Table 1

The Distribution of The Participants According To Their Departments of Current Higher-Education.

Departments	n	%
Mathematics	2	0.3
Computer	15	2.0
Foreign Language	14	1.8
Visual Arts	2	0.3
Turkish Language (Literature)	1	0.1
Law	62	8.1
Agriculture	5	0.7
Communication	40	5.2
Engineering	145	19.0
Political-Financial Sciences	243	31.8
Physical Sciences	4	0.5
Medicine	127	16.6
Computer Engineering	52	6.8
Physics/Chemistry/Biology	52	6.8
Total	764	100.0
Missing	48	

When departments of current higher-education of the participants are analyzed, the result is as follows: 243 (31,8%) from political sciences, 145 (19,0 %) from engineering and 127 (16,6 %) from medicine. It is seen that the least participation comes from mathematics, visual arts, Turkish Language (literature), agriculture and physical sciences.

Research Instrument

Occupational Field Interest Inventory (Mesleki Alan İlgi Envanteri/MAİ-OFII), developed by Deniz (2009), which aims to determine vocational interest of individuals, was used in this study.

Occupational Field Interest Inventory (OFII). This inventory aims to measure the interest of individuals in 14 vocational fields; these are Mathematics, Computer, Foreign Languages, Visual Arts, Psychology, Education, Turkish Language, Law, Agriculture-Outdoor, Communication-Mass Media, Engineering, Political-Financial Sciences, Science, Health Sciences. After reading the items, the participant marks, with reference to her/his interest, the most suitable choice from "Really attracts my attention (5)" to "Does not really attract my attention (1)". Within the scope of this study, the form of OFII consisting of 156 items, and among these 156 items the OFII short form (OFII-SF) consisting of 72 items that is formed by the items contributing most to the extend they belong, were used. Statics were calculated after raw scores were converted into z score, since the dimensions of the applied scales were same, but the numbers of items were different.

The Validity and Reliability of Research Instrument.

In some of the scale development and adaptation studies only confirmatory factor analysis (Kocayörük, 2010) was used in order to determine validity while in some others exploratory and confirmatory factor analyses (Aypay, 2012; Baloğlu, Karadağ, & Karaman, 2008; Büyüköztürk, Akgün, Özkahveci, & Demirel, 2004; Doğan & Çetin, 2009; Eren-Gümüş, 2010; Kapıkıran & Kapıkıran, 2011; Özyürek, 2010; Talepasand, Alijani, & Bigdeli, 2010; Yıldız, Akpınar, Tatar, & Ergin, 2009) were performed.

During the development of OFII, exploratory factor analysis was carried out with regard to constructive validity. As a result of this analysis it was concluded that 14 factors represent 49% of the total variance. Via confirmatory factor analysis which was carried out later on, it was seen that goodness of fit indexes obtained values between 0.87 - 0.99. In addition, for the validity studies of the scale, totally 88 academicians were asked to give their opinion about the reflection of items on the fields to be at least three academicians with PhD level in relevant field for each of 14 different fields. As another proof of the constructive validity, the correlation among 14 dimensions of the scale was calculated. The calculated correlation values were - 0.43 to 0.50 and correlation median was $r=0.07$. This result was interpreted that the dimensions were clearly separated from each other.

In some of the scale development and adaptation studies Cronbach alpha coefficients were calculated (Akin, 2008; Aypay, 2011; Baloğlu vd., 2008; Büyüköztürk vd., 2004; Yıldız vd., 2009). With regard to reliability studies of OFII, internal consistency reliability (Cronbach alpha) was ascertained and it was observed that Cronbach alpha value, which is estimated for each dimension, varied between 0.79 (Agriculture-Outdoor) and 0.95 (Law) and the median value of the reliability was 0.89. As another proof of the reliability, after form OFII consisting of 156 items was created, the test-retest reliability was estimated again and it was seen that the obtained reliability values varied from 0.75 (Agriculture-Outdoor) to 0.95 (Law) values and median value was 0.89.

Validity and reliability results gained for two forms of scale within the scope of this study are as follows: As a result of the confirmatory factor analysis; the goodness of fit indexes of the sub-dimensions involved in OFII short form were analyzed. Accordingly, it was observed that GFI (0.94-1.00), AGFI (0.82-0.99), CFI (0.89-1.00), NNFI (0.78-1.00) goodness of fit indexes and RMSEA (0-0.17) nonconformity indexes had mostly sufficient values. Besides, it was seen that when one item from each of Mathematics, Visual Arts and Foreign Language sub-dimensions was omitted, which increases the chi-square values; GFI (0.96-1.00), AGFI (0.92-0.99), CFI (0.95-1.00), NNFI (0.91-1.00) goodness of fit indexes increased and RMSEA (0-0.13) nonconformity indexes decreased.

The goodness of fit indexes on OFII form were analyzed either. Accordingly, it was observed that GFI (0.84-0.95), AGFI (0.77-0.93), CFI (0.85- 0.98), NNFI (0.81-0.98) goodness of fit indexes and RMSEA (0.063-0.14) nonconformity index had mostly sufficient values. Besides, it was seen that when one item from each of Mathematics, Psychology, Turkish Language, Communication-Mass Media, Engineering, Visual Arts and Foreign Language dimensions was omitted, which increases the chi-square

values; GFI (0.89-0.95), AGFI (0.83-0.93), CFI (0.90- 0.98), NNFI (0.87-0.98) goodness of fit indexes increased and RMSEA (0.063-0.12) nonconformity indexes decreased.

With regard to reliability study of the scale, Cronbach alpha reliability coefficients were estimated. Table 2 gives the Cronbach alpha reliability coefficients regarding OFII-SF and OFII which were estimated to the participants.

Table 2

Cronbach alpha Reliability Coefficients Regarding OFII-SF and OFII

<i>Sub-dimensions</i>	<i>OFII-SF</i>	<i>OFII</i>
Mathematics	0.76	0.88
Computer	0.86	0.93
Foreign Language	0.70	0.82
Visual Arts	0.74	0.81
Psychology	0.78	0.87
Education	0.74	0.79
Turkish Language (Literature)	0.59	0.81
Law	0.92	0.94
Agriculture-outdoor	0.67	0.76
Communication-Mass Media	0.73	0.87
Engineering	0.89	0.90
Political-Financial Sciences	0.90	0.92
Science	0.78	0.87
Health	0.86	0.89

The internal consistency values (Cronbach alpha) of OFII-SF and OFII regarding this study were found. It was observed that the Cronbach alpha value, which was estimated with respect to all dimensions regarding OFII-SF, varied between 0.59 (Turkish Language) and 0.92 (Law). It was observed that the Cronbach alpha value, which was estimated with respect to all dimensions regarding OFII, varied between 0.76 (Agriculture-outdoor) and 0.94 (Law).

Procedure

OFII and OFII-SF were applied to preparatory classes of Ankara University and Yıldırım Beyazıt University. The application was performed in 2011-2012 academic year's spring term by researchers entering the classes and giving information about the application. Within the personal information part added to the beginning of OFII and OFII-SF, the participants were asked about their departments of current higher-education, the number of books purchased, the frequency of the participation of parents in social activities, educational level of mother and father, vocation of mother and father, monthly income, possession of house and car. The SES of the participants were determined by number of books purchased, the frequency of the participation of parents in social activities, educational and income level of mother and father. Before the application, within the framework the legal rights knowledge of the participants was ensured and the motivation of the participants was improved by stating that the results regarding 14 fields of interest would be submitted. At the end

of the study, the results and their interpretations were submitted to each student in coordination with directors of both universities.

Data Analysis

After collecting the research data, the data of the ones, who filled the scale fully and correctly, were saved to computer and arranged. After the reliability and validity values of the scales were determined, the phase of data analysis started.

Before the analyzes, the raw scores obtained via OFII and OFII-SF were converted into z score. In addition to the determination of descriptive statistics for the analysis of data; chi-square ($\alpha=.05$), independent samples t test ($\alpha=.05$) and one way ANOVA statistics ($\alpha=.05$) were applied.

Results

After defining descriptive statistics in the research, the findings were respectively transferred with reference to gender and SES. Table 3 shows chi-square values of the departments of current higher-education of the students with regard to gender.

Table 3

Chi-Square Values of the Departments of Current Higher-Education of the Students with Regard to Gender

		Female(N=382)	Male(N=371)	Total
For.Lang.	n	8	6	14
	%	57	43	100
Law	n	48	14	62
	%	77	23	100
Commun	n	21	19	40
	%	52	48	100
Engine	n	50	95	145
	%	34	66	100
Political	n	113	129	242
	%	47	53	100
Medicine	n	77	50	127
	%	61	39	100
Comp.Eng.	n	32	35	67
	%	48	52	100
Physics Science Biology (P.C.B)	n	33	23	56
	%	59	41	100

$$\chi^2_{(7)}=41.56; p<.001$$

It was observed that there is significant difference with respect to gender among the departments of current higher-education of the students ($\chi^2_{(7)}=41.56; p<.001$). It can be observed that these differences are in favor of women especially in law and medicine, and in favor of men in engineering.

In the research, it was examined whether sub-dimensions of OFII showed significant difference regarding gender or not. After the z scores of the individuals were calculated, t test analysis of the OFII sub-dimensions with respect to gender was carried out. Table 4 shows these findings.

Table 4

The Results of Independent Samples t Test-to OFII Sub-Dimensions-according to Gender

	Gender	N	\bar{X}	Sx	df	t	p
Mathematics	Female	431	-0.30	1.00	809	1.66	.096
	Male	380	-0.17	1.03			
Computer	Female	431	-0.60	0.81	699.87	8.51	.000***
	Male	380	-0.02	1.07			
Foreign Language	Female	431	-0.22	1.10	809	2.29	.022*
	Male	380	-0.39	1.02			
Visual Arts	Female	431	-0.16	1.15	802.65	2.15	.031*
	Male	380	-0.32	0.92			
Psychology	Female	431	0.07	1.09	804.25	7.80	.000***
	Male	380	-0.47	0.89			
Education	Female	431	-0.29	1.07	809	1.31	.190
	Male	380	-0.39	1.00			
Turkish Language (Literature)	Female	431	-0.40	1.03	809	1.34	.178
	Male	380	-0.50	0.96			
Law	Female	431	-0.24	1.09	808.88	3.16	.002**
	Male	380	-0.47	0.97			
Agriculture-outdoor	Female	431	-0.52	0.83	775.07	2.35	.019*
	Male	380	-0.37	0.91			
Communication- Mass Media	Female	431	-0.21	1.09	809	0.56	.573
	Male	380	-0.25	0.98			
Engineering	Female	431	-0.53	1.01	772.80	3.28	.001**
	Male	380	-0.28	1.10			
Political-Financial Sciences	Female	431	-0.35	0.96	761.39	4.14	.000***
	Male	380	-0.04	1.09			
Science	Female	431	-0.35	1.03	809	2.35	.019*
	Male	380	-0.18	1.05			
Health	Female	431	-0.10	1.04	808.99	4.40	.000***
	Male	380	-0.41	0.92			

(*p<.05, **p<.01, ***p<.001)

It is observed that there are significant differences between gender and the fields of interest of the students. When Table 4 is analyzed, it is observed that in respect to the field of interests of the individuals, there is a significant difference that Psychology, Health (p<.001), Law (p<.01), Foreign Languages and Visual Arts

($p < .05$) are in favor of female; on the other hand Computer, Political-Financial Sciences ($p < .001$), Engineering ($p < .01$), Agriculture-Outdoor and Science ($p < .05$) are in favor of male.

In this study, it was examined whether the choice of departments of current higher-education showed difference with regard to SES or not. Table 5 shows the chi-square values of the departments of current higher-education the students with regard to SES.

Table 5

The chi-square Results of the departments of current higher-education of the students with regard to SES

N=538	Law		Commun		Engine.		Pol. and Fin. Sci.		Medicine		Comp Eng.		P.C.B.		
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	
Higher	13	25	10	34	23	17	48	26	29	31			8	2	2
Med.	17	33	10	34	44	46	63	35	30	32	18	36	18	49	
Low	22	42	9	32	28	37	71	39	34	37	23	46	7	19	
Total	52	100	29	100	95	100	182	100	93	100	50	100	37	100	

$\chi^2(12)=14.97; p > .05$

In Table 5, there are some differences at percentages. But these are not conspicuous. When the departments of current higher-education with regard to SES is analyzed, it is observed that there is no significant difference ($\chi^2(12)=14.97; p > .05$). It was determined that some cells' Chi-square value for the expected frequency of less than 5 was higher than 20%. As a result, the items with the lower values were excluded from further analysis.

One-way ANOVA was performed in order to determine whether SES variables (low, medium, high) would create a significant difference on the professional interest variable.

Table 6*The ANOVA Results of OFII Sub-dimensions according to SES*

		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Mathematics	Between Groups	7.57	2	3.79	3.67	0.03
	Within Groups	609.54	591	1.03		
	Total	617.11	593			
Computer	Between Groups	0.65	2	0.33	0.33	0.72
	Within Groups	588.00	591	1.00		
	Total	588.65	593			
Foreign Language	Between Groups	0.58	2	0.29	0.25	0.78
	Within Groups	695.58	591	1.18		
	Total	696.16	593			
Visual Arts	Between Groups	4.79	2	2.40	2.14	0.12
	Within Groups	662.17	591	1.12		
	Total	666.96	593			
Psychology	Between Groups	0.30	2	0.15	0.13	0.88
	Within Groups	662.83	591	1.12		
	Total	663.13	593			
Education	Between Groups	3.74	2	1.87	1.59	0.21
	Within Groups	696.02	591	1.18		
	Total	699.76	593			
Turkish Language (Literature)	Between Groups	1.76	2	0.88	0.83	0.44
	Within Groups	627.29	591	1.06		
	Total	629.04	593			
Law	Between Groups	2.59	2	1.30	1.19	0.30
	Within Groups	641.89	591	1.09		
	Total	644.48	593			
Agriculture-outdoor	Between Groups	1.31	2	0.65	0.88	0.42
	Within Groups	439.77	591	0.74		
	Total	441.08	593			
Communication -Mass Media	Between Groups	2.51	2	1.26	1.08	0.34
	Within Groups	689.69	591	1.17		
	Total	692.21	593			

Table Continued....

Engineering	Between Groups	4.13	2	2.07	1.87	0.16
	Within Groups	652.59	591	1.10		
	Total	656.72	593			
Political-Financial Sciences	Between Groups	0.69	2	0.35	0.32	0.73
	Within Groups	650.36	591	1.10		
	Total	651.05	593			
Science	Between Groups	3.24	2	1.62	1.49	0.23
	Within Groups	641.76	591	1.09		
	Total	645.00	593			
Health	Between Groups	1.42	2	0.71	0.75	0.48
	Within Groups	561.87	591	0.95		
	Total	563.291	593			

It was observed that there is no significant different for all sub-dimensions of OFII excluding Mathematics between SES and the field in which students are interested according to OFII ($F(2, 591)=3.67$; $p<.05$). As for Mathematics field, a significant difference in favor of common group between higher and medium SES was observed ($p<.05$).

Discussiona and Conclusions

In this study, the relation of the gender and SES with vocational interest was analyzed. To indicate vocational interest "OFII" was used. When the findings are examined, it is observed that the number of the ones preferring the departments at universities is in favor of women for law and medicine; in favor of men in engineering. In addition to this finding, it was found significant that the point differences with regard to OFII sub-dimensions are in favor of women for Psychology, Health, Law, Foreign Languages and Visual Art; in favor of men for Computer, Political and Financial Sciences, Engineering, Agriculture-outdoor and Physical Science. It can be asserted that these findings are in conformity with the gender-vocational interest studies. Many studies show that the interest of the women for social vocations, which require working with people, is higher, while interest of men for the vocations which require working with objects and abstract concepts is higher (Lippa, 1998, 2005; Low et al., 2005; Rounds, 1995; Su et al., 2009; Tay et al., 2009; Yılmaz, 2006). In addition, according to Low and Rounds (2007) women and men tend to be interested in activities and vocations which are stereotyped for their gender.

When above mentioned details are analyzed in detail, it is observed that the interest of the women is higher for the vocations which needs working with people (Psychology, Health, Law), and interest of men is higher for the vocation which need working with objects (Engineering, Agriculture-Outdoor, Science, Computer).

Although foreign language department is not regarded as a social vocation, the individuals graduating from this department tend to be a teacher. It corroborates the emphasis of the Low and Round (2006) about stereotyped vocations and activities for genders that women are more interested in Visual Arts. Although the political way of the Political and Financial Sciences, which is in favor of men, requires working with people, politics is an vocation imputed to men by society. The fact that there is no significant difference observed at some fields may match up with the study by Anderson et al. (1997) which indicates that there is no significant difference regarding vocational interest between women and men. Besides, certain fields such as communication and media are preferred by both men and women. This result was consistent with Çınar's (2013) finding in which the variable of gender exhibited no effect on the career and/or occupational preferences of college students and graduates.

The obtained findings regarding gender correspond to studies on personality types in Holland's theory. For instance, Deng et al. (2007) recommended realistic vocations for men and social vocations for women. In view of Holland's theory, people with realistic personality are interested in object-oriented vocations. The best examples in this manner would be vocations like engineering (Choices Interest Profiler and UCF Majors, 2012). According to Armstrong and Rounds (2008), in labor market, men tend to be dominant in investigative and realistic vocations, whereas women tend to be dominant in social and conventional vocations. All these studies explain the significant differences arising with regard to gender.

There was no significant difference observed regarding vocational interest, when the study findings were analyzed with respect to SES. In contrast, a significant difference in favor medium SES was found in mathematics interest. When the literature was reviewed, it was observed that SES is an important variable affecting vocational interest, however within the studies analyzed there was no information about which SES makes difference in which field (Betsworth & Fouad, 1997; Lent, 2005; Patton & McMahon, 2006). Within this study the number of books purchased, the frequency of the participation of parents in social activities, educational and income level of mother and father were used in determining SES. SES is a very relative concept featuring many variables. For the studies which will be made in the future, some other variables can also be preferred to determine SES. In addition, this study was performed in Ankara University and Yıldırım Beyazıt University. Since these are state universities, we may not have reached the high SES (especially for income variable). Private universities can also be included in future studies.

The results of the study show that gender is an important variable for vocational interest. Consultants who provide vocational guidance services should consider gender variable for vocational interests. It should be remembered that women can be happier with the vocations requiring to work with people and men can be happier with the vocation requiring to work with objects.

References

- Akın, A. (2008). Psikolojik iyi olma ölçekleri: Geçerlik ve güvenilirlik çalışması [The Scales of Psychological Well-being: A Study of Validity and Reliability]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 8 (3), 721-750.
- Anderson, M. Z., Tracey, T. J. G., & Rounds, J. (1997). Examining the invariance of Holland's vocational interest model across gender. *Journal of Vocational Behavior*, 50, 349 -364.
- Armstrong, P. I., & Rounds, J., (2008). Vocational psychology and individual differences. In S. Brown & R. Lent (Eds.), *Handbook of Counseling Psychology (4th ed., pp. 375-391)*. New York: Wiley.
- Aypay, A. (2011). İlköğretim II. Kademe Öğrencileri İçin Okul Tükenmişliği Ölçeği: Geçerlik ve güvenilirlik çalışması [Elementary School Student Burnout Scale for Grades 6-8: A Study of Validity and Reliability]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 11 (1), 511-527.
- Aypay, A. (2012). Ortaöğretim Öğrencileri İçin Okul Tükenmişliği Ölçeği (OOTÖ) [Secondary School Burnout Scale (SSBS)]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 12 (2).
- Balkıs, M. (2004). Kendini Araştırma Ölçeği'nin (KAÖ) Türk kültürüne uyarlanması [An adaptation study of the self-directed search in Turkish culture]. *Eğitim Araştırmaları Dergisi*, 17, 54-63.
- Baloğlu, N., Karadağ, E., & Karaman, H. (2008). Stratejik Planlama Tutum Ölçeği: Açımlayıcı ve doğrulayıcı faktör analizi çalışması [The Strategic Planning Attitude Scale: A Study of Exploratory and Confirmatory Factor Analyses]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 8 (2).
- Barak, A., & Cohen, L. (2002). Empirical examination of an online version of the self-directed search. *Journal of Career Assessment*, 10 (4), 387-400.
- Betsworth, D. G., & Fouad, N. A. (1997). Vocational interests: A look at the past 70 years and a glance at the future. *The Career Development Quarterly*. Volume 46, Issue 1, pages 23-47. DOI: 10.1002/j.2161-0045.1997.tb00689.x Article first published online: 23 Dec 2011.
- Bütün, A. (1998). *Farklı sosyo-ekonomik düzeylerdeki genel liselere ve kız meslek liselerine devam eden kız öğrencilerin mesleki ilgilerinin incelenmesi* [A Study on the vocational interest of the girls attending general and vocational high schools from various]. Unpublished doctoral dissertation. Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Büyüköztürk, Ş., Akgün, Ö., Kahveci, Ö., & Demirel, F. (2004). Güdülenme ve Öğrenme Stratejileri Ölçeği'nin Türkçe formunun geçerlik ve güvenilirlik çalışması [The Validity and Reliability Study of the Turkish Version of the Motivated Strategies for Learning Questionnaire]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 4 (2), 207-239.

- Choices Interest Profiler and UCF Majors (2012). University of Central Florida Career Services, Retrieved from http://www.career.ucf.edu/categories/Students/Undergraduate_Students/Explore_Majors_and_Careers/Choices_Interest_Profiler_and_UCF_Majors/9_0_47.aspx at 22.05.2012.
- Çınar, Y. (2013). Kariyer tercihi probleminin yapısal bir modeli ve riske karşı tutumlar: Olasılıklı DEMATEL Yöntemi temelli bütünlük bir yaklaşım [A Structural model of career preference problem and attitudes towards risk: A hybrid approach based on stochastic DEMATEL method]. *Sosyoekonomi*, 19 (1), 157-185.
- Deng, C-P., Armstrong, P. I., & Rounds, J. (2007). The fit of Holland's RIASEC model to US occupations. *Journal of Vocational Behavior*, 71, 1-21.
- Deniz, K. Z. (2009). Occupational Field Interest Inventory (OFII) development study. *Y.Y. Üniversitesi, Eğitim Fakültesi Dergisi*. Cilt:VI, Sayı:1, 289-310 <http://efdergi.yyu.edu.tr>
- Doğan, T., & Çetin, B. (2009). Tromso Sosyal Zeka Ölçeği Türkçe formunun faktör yapısı: Geçerlik ve güvenirlik çalışması [The Validity, Reliability and Factorial Structure of the Turkish Version of the Tromso Social Intelligence Scale]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 9 (2), 691-720.
- Erdil, Z. (2006). *Hacettepe üniversitesi hazırlık sınıfı öğrencilerinin mesleki ilgi ve yetenek alanları ile meslek tercihlerini etkileyen bazı faktörleri belirlenmesi* [Identification of some factors influencing the career choices and career interests and capabilities of Hacettepe University preparatory students]. Unpublished master's thesis. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Çocuk Gelişimi ve Eğitimi Programı, Ankara.
- Eren-Gümüş, A. (2010). The construct validity, reliability of Self Perception Profile for Adolescents: Original versus revised version. *Eurasian Journal of Educational Research* 39, 127-144.
- Gasser, C. E., Larson, L. M., & Borgen, F. H. (2007). Concurrent validity of the 2005 Strong Interest Inventory: An examination of gender and major field of study. *Journal of Career Assessment*, 15 (1), 23-43.
- Helledey, K. I., Zytowski, D. G., & Fouada, N. A. (2004). Kuder Career Search: Test-retest reliability and consequential validity. *Journal of Career Assessment*, 12 (3), 285-297.
- Herr, E.L., & Cramer, S. H. (1996). *Career guidance and counseling through the lifespan* (Fifth Edition). New York: Longman Inc.
- Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments*. (2. Edition). USA: Prentice Hall.
- International Association for Educational and Vocational Guidance. (2002). Declaration of educational and vocational guidance. *International Journal for Educational and Vocational Guidance*, 2, 79-83.

- Kapıkıran-Acun, N., & Kapıkıran, Ş. (2011). School climate inventory: Exploratory and confirmatory factor analysis and reliability-validity. *Eurasian Journal of Educational Research*, 42, 117-134.
- Karasar, N. (2011). *Bilimsel araştırma yöntemi* [Scientific research methodology]. Ankara: Nobel
- Kocayörük, E. (2010). A Turkish adaptation of the inventory of parent and peer attachment: The reliability and validity study. *Eurasian Journal of Educational Research*, 40, 131-153.
- Kuzgun, Y. (2000). *Meslek danışmanlığı* [Career counseling]. Ankara: Nobel
- Külahoğlu, Ş. (2001). Meslek gelişimine bütünsel bir bakış: Gottfredson'un kuramı [An integral view in vocational development: Gottfredson's theory]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 12, 1,
- Lent, R. W. (2005). A social cognitive view of career development and counseling. In S.D. Brown & R.W. Lent (Eds.), *Career development and counseling: Putting theory and research to work*. (pp. 101-130). New York: Wiley.
- Lippa, R. A. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology*, 74, 996-1009.
- Lippa, R. A. (2005). Subdomains of gender-related vocational interests: Do they form a cohesive bipolar M-F dimension? *Journal of Personality*, 73, 693-729.
- London, H. J. E., Würzburg, W. A., & Berne, R. M. (Eds.), (1972). *Encyclopedia of psychology*. New York, USA: The Seabury.
- Low, K. S. D., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: A quantitative review of longitudinal studies. *Psychological Bulletin*, 131, 713-737.
- Low, K. S. D. & Rounds, J. (2006). Vocational interests: Bridging person and environment. In D. L. Segal & J. Thomas (Eds.), *Comprehensive handbook of personality and psychopathology, Volume I: Personality and everyday functioning* (pp. 251-267). NY: Wiley.
- Low, K. S. D., & Rounds, J. (2007). Interest change and continuity from early adolescence to middle adulthood. *International Journal of Educational and Vocational Guidance*, 7, 23-36.
- Lubinski, D. & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, 1, 316-345.
- Özyürek, R. (2010). Matematik Yetkinlik Beklentisi Bilgilendirici Kaynaklar Ölçeği'nin geçerlik ve güvenilirlik çalışmaları [The Reliability and Validity of the Mathematics Self-Efficacy Informative Sources Scale]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 10 (1).
- Patton, W., & McMahon, M. (2006). *Career development and systems theory: Connecting theory and practice* (2nd ed.). Rotterdam, The Netherlands: Sense.

- Rottinghaus, P. J., Coon, K. L., Gaffey, A. R., & Zytowski, D. G. (2007). Thirty-year stability and predictive validity of vocational interests. *Journal of Career Assessment*, 15 (1), 5-22.
- Rounds, J. B. (1995). Vocational interests: Evaluation of structural hypotheses. In D. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 177-232). Palo Alto, CA: Consulting Psychologists Press.
- Rounds, J., & Day, S. X. (1999). Describing, evaluating, and creating vocational interest structures. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Their meaning, measurement and use in counseling* (pp. 103-133). Palo Alto, CA: Davies-Black.
- Savickas, M. L. (1999). The psychology of interests. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning, measurement and counseling use*. USA: Davies-Black.
- Sayın, S. (2000). *Lise öğrencilerinin mesleki ilgilerini yordayan bazı değişkenler* [Some variables predicting high school students vocational interest]. Unpublished doctoral dissertation. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Silvia, P. J. (2006). *Exploring the psychology of interest*. New York: Oxford University Press.
- Strong, E. K. (1943). *Vocational interests of men and women*. Stanford, California: Stanford University.
- Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135, 859-884.
- Talepasand, S., Alijani, F., & Bigdeli, I. (2010). Validation of the Social Achievement Goal Orientation Scale in Iranian students. *Eurasian Journal of Educational Research*, 40, 17-31.
- Tay, L., Drasgow, F., Rounds, J., & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal? *Journal of Applied Psychology*, 94, 1287-1304.
- Varçın, R., Ergün, G., Gülçubuk, B., Öğülmüş, S., Özgenç, Ö.Y., Pişkin, M., Savcı, İ., & Deniz, K. Z. (2005). *İşletmelerde kariyer planlaması el kitabı*, [The handbook of career planning in Business]. (Ed. R. Varçın). Ankara: Odak.
- Yıldız, E., Akpınar, E., Tatar, N., & Ergin, Ö. (2009). İlköğretim öğrencileri için geliştirilen Bilgi Üstü Ölçeği'nin açımlayıcı ve doğrulayıcı faktör analizi [Exploratory and Confirmatory Factor Analysis of the Metacognition Scale for Primary School Students]. *Kuram ve Uygulamada Eğitim Bilimleri Dergisi*, 9 (3), 1573-1609.
- Yılmaz, S. (2006). Fen Bilgisi öğretmen adaylarının mekanik konularına karşı tutumlarının incelenmesi [Analysis of Preservice Science Teachers' Attitudes toward Mechanics' Concepts]. *Eğitim Araştırmaları Dergisi*, 24, 199-208.

Cinsiyet ve Sosyo-ekonomik Düzey Açısından İncelenmesi

Özet

Atf:

Deniz, K. Z., Türe, E., Uysal, A., & Akar, T. (2014). Investigation of vocational interest and vocational preference in terms of gender and socio-economic status. *Eurasian Journal of Educational Research*, 57, 91-112
<http://dx.doi.org/10.14689/ejer.2014.57.1>

Problem Durumu

Bireyler yaşamlarını sürdürme, beslenme, barınma, güvencede olma, toplumda iyi bir konumda olma vb. amaçları gerçekleştirmek için bir mesleği tercih etme eğilimindedir. Genellikle ergenlik dönemi görevlerinden birisi olan meslek seçimi kimine göre yaşamın en önemli adımı, kimine göre ise yaşamda çok önemli ama her an değiştirilebilir ve esnetilebilir bir adımdır. Meslek seçmek bazen yaşam biçimini seçmektir. Mesleki tercih, bir grup meslekten birini seçme süreci olarak tanımlanmaktadır. Birey mesleki tercih yaparken genellikle kendisi için uygun olan ve olmayanları belirlemeye yönelik bir süreç geçirir. Bireyin hedeflerinin belirlenmesinde, değerlerinin, ilgilerinin ve kariyer kararlarının sürekli gözden geçirilmesi gerektiği vurgulanmaktadır.

Cinsiyet ilgileri ayırmada önemli bir değişken olarak karşımıza çıkmaktadır. Yapılan pek çok araştırma, literatür taraması ve meta analizi çalışması erkeklerin nesnelere kadınlarinsa insanlarla birlikte çalışmaya ilişkin mesleklerle ilgilendiklerini göstermektedir. Aynı zamanda tercihleri de ilgilerine paralel olarak bu şekilde ortaya çıkmaktadır. Mesleki ilgilerin belirlenmesinde bir başka değişken sosyo-ekonomik düzey (SED)'dir. Yapılan araştırmalar meslek seçimi ve mesleki ilginin cinsiyet, etnik köken, genetik, fiziksel sağlıkla birlikte SED gibi pek çok değişkenden etkilendiğini göstermektedir. Bireylerin mesleki ilgileri ve mesleki tercihleri etkileyen faktörleri ortaya koymak mesleki yönlendirmede kararlarında daha isabetli olmasını sağlayacaktır.

Araştırmanın Amacı

Bu araştırmanın amacı Mesleki Alan İlgi Envanteri (MAİ) yardımıyla mesleki ilgileri ve mesleki tercihleri etkileyen faktörleri ortaya koymaktır.

Araştırmanın Yöntemi

Araştırmaya 2011-2012 eğitim-öğretim yılı bahar döneminde Ankara Üniversitesi'nin (N=495) ve Yıldırım Beyazıt Üniversitesi'nin (N=317) hazırlık sınıflarında öğrenim gören 812 öğrenci katılmıştır. Bunlardan 243'ü siyaset bilimleri, 145'i mühendislik 127'si tıp bölümlerinde öğrenim görmektedir. En az katılımcı sayısının ise matematik, görsel sanatlar, dil (edebiyat), ziraat ve fen bilimleri bölümlerindedir.

Araştırmada bireylerin mesleki ilgilerini tespit etmeyi amaçlayan Mesleki Alan İlgi Envanteri'nin uzun (MAİ) ve kısa formu (MAİ-KF) kullanılmıştır. Araştırmada bireylerin mesleki ilgilerini tespit etmeyi amaçlayan MAİ'nin 156 maddelik formu ve

bu 156 madde içinden ait olduğu boyuta en çok katkısı yapan maddelerden oluşturulan 72 maddelik MAİ kısa formu (MAİ-KF) kullanılmıştır.

MAİ'nin geliştirme çalışmasında yapı geçerliğine yönelik olarak önce açılımlayıcı faktör analizi yapılmıştır. Bu analiz sonucunda 14 faktörün toplam varyansın %49'unu açıkladığı sonucuna ulaşılmıştır. Daha sonra yapılan doğrulayıcı faktör analizi sonucunda uyum indekslerinin 0,87 ile 0,99 arasında değer aldığı görülmüştür. Bunun yanı sıra ölçeğin geçerlik çalışmalarına yönelik olarak 14 farklı alanın her birinden alanında en az doktora derecesine sahip en az üç akademisyen olmak üzere toplamda 88 akademisyenden maddelerin alanlarını yansıtmaları hakkında görüş alınmıştır. Yapı geçerliğinin bir başka kanıtı olarak ölçeğin 14 boyutunun kendi aralarındaki korelasyonları hesaplanmıştır. Hesaplanan korelasyon değerleri -0,43 ile 0,50 arasında değer almıştır ve korelasyonların medyanı $r=0,07$ 'dir.

MAİ'nin güvenilirlik çalışmalarına yönelik olarak iç tutarlılık güvenilirliği (Cronbach alfa) tespit edilmiştir ve her boyut için hesaplanan Cronbach alfa değerinin 0,79 (Ziraat) ile 0,95 (Hukuk) arasında değiştiği, güvenilirliklerin medyan değerinin ise 0,89 olduğu gözlenmiştir. Güvenirlik için bir başka kanıt olarak MAİ 156 maddelik form oluşturulduktan sonra test tekrar test güvenilirlik değeri hesaplanmış ve elde edilen güvenilirlik değerlerinin 0,75 (Ziraat) ile 0,95 (Hukuk) arasında değiştiği medyan değerinin ise 0,89 olduğu görülmüştür.

Analizler yapılmadan önce MAİ ve MAİ-KF'den elde edilen ham puanlar z puanına dönüştürülmüştür. Verilerin analizinde betimsel istatistiklerin tespit edilmesine ek olarak, ki-kare, t testi ve tek yönlü ANOVA istatistiksel teknikleri kullanılmıştır.

Araştırmanın Bulguları

Araştırma sonucunda bireylerin okudukları bölümler ve ilgi alanları açısından cinsiyete göre anlamlı bir farklılık olduğu gözlenmiştir. Bireylerin öğrenim gördükleri bölümler arasındaki farklılıkların özellikle hukuk ve tıp alanlarında kadınlar lehine, mühendislik alanında erkekler lehine olduğu; ilgi alanları bakımından Psikoloji, Sağlık, Hukuk, Yabancı Dil ve Görsel Sanatlar alanlarında kadınlar lehine; Bilgisayar, Siyasal-Mali Bilimler, Mühendislik, Ziraat ve Fen Bilimleri alanlarında ise erkekler lehinedir.

SED'e göre elde edilen bulgularda ise, bireylerin bölüm tercihleri ve ilgi duydukları alanlar (matematik hariç) açısından anlamlı bir farklılık olmadığı sonucuna ulaşılmıştır. Matematik alanında ise üst ile orta SED arasında orta grup lehine anlamlı fark gözlenmiştir.

Araştırmanın Sonuçları ve Önerileri

Araştırma bulgularının literatürdeki cinsiyet-mesleki ilgi/tercih çalışmalarıyla uyumlu olduğu söylenebilir. Yapılan pek çok araştırma, insanlarla çalışma gerektiren sosyal ağırlıklı mesleklerde kadınların; nesne ve soyut kavramlarla çalışma gerektiren mesleklerde ise erkeklerin ilgisinin/tercihinin daha yüksek olduğunu göstermektedir. Kadınlar ve erkekler kendi cinsiyeti için kalıplaşmış etkinlikler ve mesleklere ilgi duyma eğilimindedirler.

Kadınların ağırlıklı olarak insanlarla (Psikoloji, Sağlık, Hukuk), erkeklerin ise ağırlıklı olarak nesnelere (Mühendislik, Ziraat, Fen Bilimleri, Bilgisayar) çalışılan mesleklere ilgilerinin yüksek olduğu gözlenmektedir. Yabancı dil alanı sosyal bir

meslek gibi görülmemesine rağmen bu bölümden mezun olanların çoğunlukla öğretmenliğe yöneldiği görülmektedir. Kadınların ilgilerinin Görsel Sanatlar alanında yüksek çıkması cinsiyetler için kalıplaşmış meslek ve etkinlik vurgusunu doğrulamaktadır. Erkeklerin lehine olan Siyasal-Mali Bilimlerin siyasal boyutu insanla çalışma gerektirse de siyaset toplum tarafından erkeğe atfedilmiş bir meslek alanıdır. Bazı alanlarda anlamlı farklılık gözlenmemesi ise literatürde mesleki ilgi açısından kadınlar ve erkekler arasında anlamlı farklılık gözlenmeyen çalışmalarla örtüşebilir. Ayrıca iletişim ve medya gibi bazı sektörler hem kadın hem de erkeklerin birlikte tercih ettiği mesleki alanlardır.

Bu araştırmada SED belirlenirken eve alınan kitap sayısı, anne-babanın sosyal etkinliğe katılma sıklığı, anne-babanın eğitim düzeyi ve gelir düzeyinden yararlanılmıştır. SED kavramı oldukça göreceli ve birçok değişkeni içinde barındıran bir kavramdır. Yapılacak olan yeni çalışmalarda SED belirlemek için farklı değişkenler de tercih edilebilir.

Araştırma sonuçları cinsiyetin meslek seçiminde ve mesleki ilgilerde önemli bir değişken olduğunu göstermektedir. Mesleki rehberlik yapan danışmanlar mesleki tercihlerde cinsiyet değişkenini dikkate almalıdır. Kadınların insanlarla çalışma gerektiren mesleklerde, erkeklerinse insanlarla çalışma gerektirmeyen özellikle nesne ağırlıklı mesleklerde daha mutlu olabileceği unutulmamalıdır.

Anahtar Sözcükler: Mesleki İlgi, Mesleki Tercih, Cinsiyet ilişkisi, SED ilişkisi, Mesleki Alan İlgi Envanteri

Measuring Essay Assessment: Intra-rater and Inter-rater Reliability

Ulaş KAYAPINAR*

Suggested Citation:

Kayapınar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136
<http://dx.doi.org/10.14689/ejer.2014.57.2>

Abstract

Problem Statement: There have been many attempts to research the effective assessment of writing ability, and many proposals for how this might be done. In this sense, rater reliability plays a crucial role for making vital decisions about testees in different turning points of both educational and professional life. Intra-rater and inter-rater reliability of essay assessments made by using different assessing tools should also be discussed with the assessment processes.

Purpose of Study: The purpose of the study is to reveal possible variation or consistency in grading essay writing ability of EFL writers by the same/different raters using general impression marking (GIM), essay criteria checklist (ECC), and essay assessment scale (ESAS), and discuss rater reliability.

Methods: Quantitative and qualitative data were used to present the discussion and implications for the reliability of ratings and the consistency of the measurement results. The assessing tools were applied to 44 EFL university students and 10 graders assessed the essay writing ability of the students by using GIM, ECC, and ESAS in different occasions.

Findings and Results: The findings and results of the analyses indicated that using general impression marking is evidently not reliable for assessing essays. The coefficients obtained from checklist and scale assessments, considering the correlation coefficients, estimated variance components, and generalizability coefficients present valuable information, clearly show that there is always variation among the results.

* Dr., EPP, American University of the Middle East, Kuwait. ukayapinar@gmail.com

Conclusions and Recommendations: When the total scores and the rater consensus results in this study are examined, it can be clearly seen that the scores are almost always not identical and they are different from each other. For this reason, opposed to the idea that is commonly agreed upon, checklists or even scales may not be effectively as reliable as expected and they may not improve inter-reliability or intra-reliability of ratings unless the raters are very well-trained and they have strong agreement or common inferences on performance indicators and descriptors since they should not have ambiguous interpretations on the criteria set. The results might be more accurate and reliable if the accepted interpretation of a meaningful correlation coefficient for this kind of measurements can be considered as .90 minimum for giving evidence of reliable ratings. This might mean that the proximity of the scores which are assigned to same or independent essays will be higher and more similar. However, the scale use could still be emphasized as more reliable. Still, an elaborate and careful examination with more raters is seen needed.

Keywords: Essay, assessment, intra-rater, inter-rater, reliability.

Assessing writing ability and the reliability of ratings have been a challenging concern for decades and there is always variation in the elements of writing preferred by raters and there are extraneous factors causing variation (Blok, 1985; Chase, 1968; Chase, 1983; Darus, 2006; East, 2009; Engelhard, 1994; Gyagenda & Engelhard, 1998a; Gyagenda & Engelhard, 1998b; Hughes, Keeling & Tuck, 1980; Hughes, Keeling & Tuck, 1983; Hughes & Keeling, 1984; Kan, 2005; Klein & Hart, 1968; Klein & Taub, 2005; Marshall & Powers, 1969; Murphy & Balzer, 1989; Schaefer, 2008; Slomp, 2012; Sulsky & Balzer, 1988; Wexley & Youtz, 1985; Woehr & Huffcutt, 1994). Fisher, Brooks, and Lewis (2002) state fitness for purpose requirement is the core of all testing work, and direct writing assessments are subjective and thereby more prone to reliability issues. For this reason, many raters use scoring scales or rubrics because they believe that any assessment without a scale is based on subjective judgments and general impression. Some researchers also state that not only general impression marking but also holistic assessment with a set of criteria can be highly subjective (Hamp-Lyons, 1991; Vaughan, 1991) and scores can vary in a significant way. Huot (1990) states that the levels of reliability achieved with holistic assessment are generally lower than that achieved with analytic assessment (Johnson, Penny, & Gordon, 2001). In this respect, general impression marking and holistic assessment can be called as subjective but analytic assessment can be called objective-like or systematically subjective because, all in all, each indicator of criteria is scored subjectively (Kayapınar, 2010). Even if it seems more reliable than the others, there is still a set of criteria which is implicit or explicit for different types of assessment. Moreover, a comparison of reliability measures by using different assessment tools is seen necessary in order to provide evidence going beyond any claim and reaching the proof of assessing essays consistently because the rating

methods –holistic or analytic- used by the raters can change their application of rating criteria (Huang, 2012).

In this article, general impression marking refers to handling with an essay as a whole with a subjective judgment (Hamp-Lyons, 1992). For this reason, no tool was addressed for this type of assessment in the study. Holistic assessment refers to scoring the overall product as a whole, with judging the predetermined component parts separately (Mertler, 2001; Nitko, 2001). For this type of assessment, a checklist entitled Essay Criteria Checklist (App.1) was employed. A rating scale entitled Essay Assessment Scale (App.2) was used for analytic assessment which refers to scoring the levels of the product with individual predetermined criteria and obtaining a total score by the sum of the individual scores (Moskal, 2000; Nitko, 2001; Weir, 1990).

Considering the measures of rater reliability and the carry-over effect, the basic research question guided in the study is in the following:

Is there any variation in intra-rater reliability and inter-reliability of the writing scores assigned to EFL essays by using general impression marking, holistic scoring, and analytic scoring?

Method

Sample

Three study groups were randomly chosen and employed as follows: *Judges*. Judges ($n=103$) include faculty of ELT departments from different (20) universities. They evaluated the appropriateness and validity of the checklist items (App. 1) and the criteria and performance indicators of the scale (App. 2). *Raters*. Raters ($n= 10$) who assessed the essays are ELT experts (MAs and PhDs) and experienced teachers of writing skill (at least 2 years). *EFL students*. The students ($n= 44$) who responded the essay test produced the essays in testing conditions for Advanced Reading and Writing class.

Research Instruments

The writing samples. Forty-four scripts of one essay sample written in testing conditions in order to achieve the objective :

“By means of the awareness of essay types, essay writers will analyze, synthesize and evaluate information and therefore, in their compositions, react to prompts. Essay writers will also be able to analyze and produce different types of essays (e.g. comparison and contrast, classification, process analysis, cause-and-effect analysis, and argumentative) that are unified, coherent, and organized.” The essay prompt, which was produced by the teachers of the particular class, is the same for all students as: *Please write an essay about the topic “University students should be free to choose their own courses.”*

Essay Criteria Checklist (ECC). The checklist was developed in order to measure each construct of essay writing. First of all, a criteria list was written through a review of relevant literature (Raimes, 1983; Norton, 1990; Celce-Murcia, 2001; Johnson, Penny, & Gordon, 2001; Jacobs *et al.* 1981 in Weigle, 2002; Weigle, 2002;

Bowen and Cali, 2004; Hawkey & Barker, 2004; Darus, 2006; IELTS, 2007; Dempsey, PytlikZillig, & Bruning, 2009; Knoch, 2009). Next, 103 faculty from ELT departments from different (20) universities examined the appropriateness of the checklist considering the expressions used and the consistency between the objectives and constructs of essay writing skill and the checklist items. The ratio of agreement (P) (Erkuş, 2003) was found significantly high ($P=.96.1$; $P = \frac{\text{the number of judges agreed on each criterion}}{\text{total number of judges}}$). Later, two experts of measurement and evaluation examined the checklist considering the content and technical features.

Essay Assessment Scale (ESAS). The scale was developed in order to describe and measure each construct of essay writing skill with performance levels. First, 103 faculty of ELT departments from different (20) universities examined the scale considering the expressions used and the consistency between the objectives and constructs of essay writing skill and the performance indicators included. The ratio of agreement (P) of the scale is also .96.1. Next, two experts of measurement and evaluation examined the scale considering technical features. Finally, a Likert type scale covering five performance levels (0-1-2-3-4) was developed by using expert judgments. Five performance levels were chosen because of easiness and usefulness for the observable behavior although there is no limit for performance levels (Kan, 2007).

The measurement results: The total scores of 2640 ((10 raters \times 44 essay scripts) \times 6 independent sessions) essay scripts, which were randomly selected, were used to measure the reliability of ratings, using GIM, ECC, and ESAS.

Standardized open-ended interviews. The raters were asked the following standardized open-ended interview questions about the assessment process:

1. "What do you think of the assessments you made by using GIM?"
2. "What do you think of the assessments you made by using ECC?"
3. "What do you think of the assessments you made by using ESAS?"

A pretest of the interview questions was carried out by two independent raters and two experts of measurement and evaluation in order to identify the validity and the effectiveness of the questions.

Procedure

The procedure of the study includes two phases: *The production of the material to be scored.* The essays were produced in testing conditions of an advanced reading and writing class. Each essay was given a different code assigned randomly for each rating after the names had been deleted.

Assessment Design. There are ten raters and six different rating processes in the study. Before the raters started each rating session, they had been given a short educational session and instructions for a proper completion of each session. Each rater scored each essay at a time -44 essays in one batch and 264 essays in total. Each rating session was held after a 10-week break in order to remove the carry-over effect

of the previous assessment. In order to balance the objectivity, the order and the numbering of the essays were changed before each session and they were assigned random codes.

Data Analyses

In order to determine the intra-rater reliability of the ratings, the correlation coefficients between the two gradings of the same raters for the same essays were computed by using Pearson Product Moments Correlation Analysis. The correlation coefficients were also examined by using Fischer's *z* Transformation to test the significance of the variation in correlation coefficients. This procedure led the way to put the correlation coefficients in order. ANOVA was employed in order to present evidence for the inter-rater reliability of ratings. The differences in the scores across the task and the raters by using GIM and ESAS were also interpreted through a generalizability study. A series of person \times rater \times task were performed to examine the variation of scores due to potential effects of person, rater, and task after the variance components had been estimated. Using standardized open-ended interviews revealed the reflections and views of the raters on their own rating process. The qualitative data here were analyzed line by line and memos were written (Glesne, 1999; Strauss & Corbin, 1998). Categories were reviewed and recurring themes, core consistencies and meanings were identified by using pattern codes. Those explanatory pattern codes were later identified as smaller sets and themes with content analysis (Miles & Huberman, 1994; Patton, 2002). The process includes: Underlying key terms in the responses, restating key phrases, coding key terms, pattern coding, constructing themes, and incorporating themes into an explanatory framework

Results

Intra-rater reliability.

Table 1 shows the intra-rater consensus between GIM assessments.

Table 1

Intra-rater Consensus between GIM Assessments

Difference	R1		R2		R3		R4		R5		R6		R7		R8		R9		R10	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
0	7	16	2	5	6	14	6	14	1	2	1	2	1	2	9	21	7	15	3	7
±1-5	9	21	17	38	8	18	18	41	13	30	30	68	10	23	7	15	1	3	18	41
±6-10	8	18	7	15	9	21	8	18	13	30	12	27	6	14	7	15	7	15	14	32
±11-15	9	21	6	14	8	18	4	9	6	14	0	0	6	14	9	21	2	5	2	5
±15- more	11	25	12	27	13	30	8	18	11	25	1	2	21	47	12	27	12	27	7	15
TOTAL	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100

R=Rater

Table 1 shows that Rater 6 scored 31 essays out of 44 with a $\pm 0-5$ -point difference on 0-100 point scale. This is the highest value among the others referring that 70% of the essays have similar results in two assessments made by using GIM. The assessments of Rater 9 have the lowest percentage of consensus which is 18% with a $\pm 0-5$ -point difference. The frequency is 7 for zero difference, and 1 for $\pm 1-5$ -point difference. Other raters' consensus between two assessments by using GIM has a frequency range between 11 and 21 points. Table 2 also indicates that the percentages of the scores which are the same in two assessments have a range between 2 and 21. This means that the frequencies range between 1 and 9 out of 44 essays. Rater 5, 6, and 7 have only one score which is the same for both assessments. However, Rater 8 scored 9 essays the same. For a better understanding of the rater reliability of general impression marking, it is necessary to examine the correlation coefficients between the two assessments made by using GIM. The correlation coefficients computed, by using Pearson Product Moments Correlation, are presented below in Table 2:

Table 2*Correlations across GIM Assessments*

Rater	<i>r</i>
1	.042
2	.510**
3	.477**
4	.279
5	.450**
6	.835**
7	.584**
8	.412**
9	.790**
10	.880**

** Correlation is significant at the 0.01 level

The correlation coefficients, seen in Table 2, range between .042 and .880. Among the ten coefficients, two of them, which belong to the raters 1 and 4, are not significant. The other correlation coefficients seem significant. This may mean that those raters assigned similar scores to the essays in both assessments. However, only 3 of them are above .70 which refers to a considerably high and meaningful correlation (Kline, 1986) and relatively a high consistency. In fact, even the coefficient of .70 seems insufficient for a high level of consistency when the intra-rater consensus is examined and the results in Table 1 and 2 are compared carefully. For example, Rater 10 scored only 3 essays (7%) with no difference and 18 essays (41 %) out of 44 with a ± 1 -5-point difference in spite of the highest correlation coefficient obtained (.880) among GIM assessments.

Table 3*Intra-rater Consensus Between ECC Assessments*

Difference	R1		R2		R3		R4		R5		R6		R7		R8		R9		R10	
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%
0	1	2	1	2	1	2	3	7	6	14	1	2	5	11	0	0	1	2	2	5
±1-5	9	21	35	80	14	32	14	32	36	82	33	75	9	21	40	91	30	69	21	48
±6-10	12	27	8	8	7	16	14	32	2	5	10	23	9	21	3	7	12	27	17	39
±11-15	6	14	0	0	10	23	11	25	0	0	0	0	11	25	0	0	1	2	4	9
±15-more	14	32	0	0	12	27	2	5	0	0	0	0	36	82	1	2	0	0	0	0
TOTAL	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100

R=Rater

Table 3 shows that Rater 5 scored 42 essays out of 44 with a ± 0 -5-point difference on 0-100 point scale although there are 6 essays scored with a zero difference. This is the highest value among the others referring that 96% of the essays have closer results to each other in two ECC assessments. The assessments of Rater 1 have the lowest percentage of consensus which is 23% with a ± 0 -5-point difference. The frequency is also 1 for zero difference, and 9 for ± 1 -5-point difference. Other raters' consensus between two assessments by using ECC has a frequency range between 15 and 40 points. Table 4 also indicates that the percentages of the scores which are the same in two assessments have a range between 2 and 14. This means that the frequencies range between 1 and 6 out of 44 essays. Rater 8 has no score which is the same for two assessments and the raters 1, 2, 3, 6, and 9 have only one score which is the same for two assessments. However, Rater 5 scored 6 essays the same. For a better understanding, it is necessary to examine the correlation coefficients between the two assessments made by using ECC. The correlation coefficients computed, by using Pearson Product Moments Correlation, are presented below in Table 4:

Table 4
Correlations across ECC Assessments

Rater	<i>r</i>
1	.072
2	.953**
3	.517**
4	.457
5	.955**
6	.898**
7	.730**
8	.932**
9	.928**
10	.804**

** Correlation is significant at the 0.01 level

In Table 4, the correlation coefficients range between .072 and .932, this is relatively higher than the correlation coefficients across GIM assessments. Among the ten coefficients, only one of them, which belong to the scores assigned by the rater 1, is not significant. The other correlation coefficients seem significant. This may mean that those raters gave similar scores to the essays in both assessments. However, 7 of them are above .70 which refers to a high and meaningful correlation coefficient and relatively a high consistency (Kline, 1986). Table 5 below shows the intra-rater consensus between ESAS assessments:

Table 5
Intra-rater Consensus between ESAS Assessments

<i>Difference</i>	<i>R1</i>		<i>R2</i>		<i>R3</i>		<i>R4</i>		<i>R5</i>		<i>R6</i>		<i>R7</i>		<i>R8</i>		<i>R9</i>		<i>R10</i>	
	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>
0	9	21	5	11	3	7	3	7	6	14	2	5	2	5	6	14	3	7	3	7
±1-5	18	41	21	48	18	41	3	7	28	64	18	41	24	55	24	55	23	53	24	55
±6-10	8	18	4	9	11	25	6	14	10	23	7	16	8	18	12	28	12	28	10	23
±11-15	4	9	14	32	7	16	7	16	0	0	6	14	2	5	2	5	4	9	6	14
±15-more	5	11	0	0	5	11	25	57	0	0	11	25	8	18	0	0	2	5	1	2
TOTAL	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100	44	100

R=Rater

Table 5 shows that Rater 1 scored 27 essays out of 44 with a ±0-5-point difference on 0-100 point scale. This means 62% of the essays have similar results in two assessments made by using ESAS. In the assessments of Rater 2, the number of the essays scored with ±0-5-point difference is 26, and the percentage is 59%. Rater 3 scored 21 essays with ±0-5-point difference, which means 48%. Rater 4 is the one who has the smallest amount of consistency. The rater scored only 6 essays with ±0-5-

point difference, which refers to 14%. In the assessments of Rater 5, the number of the essays scored with $\pm 0-5$ points difference is 34, which is quite high (78%) when compared to others. The results of Rater 6 show that 20 essays were scored with $\pm 0-5$ -point difference on 0-100 point scale. Rater 7 scored only 2 essays the same but there are 26 essays scored with a $\pm 0-5$ -point difference. Assessments of Rater 8 indicate 30 essays have $\pm 0-5$ -point difference which refers to 69%. In the assessments made by Rater 9, the number of essays with $\pm 0-5$ -point difference is 26. Finally, Rater 10 scored 27 essays with $\pm 0-5$ -point difference with a percentage of 62. For a better understanding of the rater reliability of the scale, it is necessary to examine the correlation coefficients between the two assessments made by using ESAS. The correlation coefficients computed, by using Pearson Product- Moment Correlation, between the first and the second assessments and they are presented below in Table 6:

Table 6*Correlations across ESAS Assessments*

Rater	<i>r</i>
1	.757**
2	.641**
3	.585**
4	.021
5	.825**
6	.680**
7	.545**
8	.916**
9	.811**
10	.884**

** Correlation is significant at the 0.01 level

The results indicate that the correlation coefficients between the scores raters assigned to the essays seem to be high and significant at the 0.01 level (no less than .545) except the one which was done by Rater 4 (.021). These results refer that 9 raters scored the essays in a significantly reliable way. Moreover, 7 of the correlation coefficients are around .70. This is a high level of positive correlation which is seen meaningful and which might mean that there is a high consistency between the assessments (Kline, 1986). When the results are compared to the others, Rater 4 is the one who has the smallest amount of intra-rater consistency, correspondingly, the one whose results have the lowest and the only insignificant correlation coefficient. The highest correlation coefficient belongs to Rater 8 (.916) whose scores correspond to each other. This refers to similar results for two assessments made in different time

distances. Moreover, Rater 8 is the one who scored 42 essays out of 44 with ± 10 points difference on 0-100 point scale (intra-rater consensus=95%). This is the best result among the raters' assessments; however, the differences among the correlation coefficients, even the ones within a 10-point difference in total scores, of the same essays scored in different times indicate there is always a source of variation in assessments made by ESAS.

Table 7

The Comparisons among Correlation Coefficients across Different Assessments

		Raters									
		1	2	3	4	5	6	7	8	9	10
The Difference between Correlation Coefficients	$r_{12} - r_{34}$	0.056	p<.05 2.433	0.099	0.016	p<.05 2.992	0.481	0.487	p<.05 2.311	1.071	0.498
	$r_{12} - r_{56}$	1.772	0.369	0.282	0.867	1.657	0.702	0.107	0.106	0.109	0.034
	$r_{34} - r_{56}$	1.648	1.572	0.176	0.849	1.282	1.137	0.141	0.205	0.961	0.531

In the table showing Fischer's z transformation, r_{12} refers to the correlation coefficient between the first two ratings; r_{34} refers to the correlation coefficient between the following two ratings; and r_{56} refers to the correlation coefficient between the final ratings. The differences at the significant level ($p < 0.05$) are presented in the table. The results indicate that few raters (2, 5, and 8) made consistent and decisive assessments in different time distances. As seen in the table, no other consistent and decisive assessments were made by the raters using the same tools in different time distances. This may mean raters assign different scores to the same essays in different time distances.

Inter-rater reliability

An analysis of variance was conducted to find out the inter-rater consensus statistically. The results are given in the table below:

Table 8*Inter-rater Reliability of Assessments*

Rating	Sum of Squares	df	Mean Square	F	Sig.
1	17554.036	9	1950.448	11.052	.000
2	21461.411	9	2384.601	8.913	.000
3	22407.909	9	2489.768	13.465	.000
4	20462.684	9	2273.632	10.164	.000
5	17570.475	9	1952.275	15.781	.000
6	31722.773	9	3524.753	31.983	.000

$p < 0.001$

The table shows the output of the ANOVA analysis and whether there is a statistically significant difference between group means. The results apparently indicate that the paired comparisons of the means of the scores raters assigned to the essays significantly differ from each other. It is clearly seen that the significance level is 0.000, which is below 0.001 ($p < 0.001$). Therefore, there is a clear statistically significant difference in the mean scores assigned by different raters. This might mean that there are remarkable differences among scores assigned by the raters to the same essay products and the inter-rater reliability of the assessments is considerably low.

A series of a random one-facet (student \times rater) model and a random two-facet model (student \times task \times rater) generalizability study for each rating (GIM and ESAS) were performed. It could not be realized for ECC ratings because of data loss. In addition, the generalizability study could be held for 9 raters as one of the raters was not able to provide the data for it as well. Estimated variance components for the ratings are given in Table 9 below:

Table 9*Estimated Variance Components (EVC) for GIM and ESAS ratings*

<i>Source</i>	<i>n</i>	<i>GIM</i>		<i>ESAS</i>	
		<i>EVC</i>	<i>Total Variance %</i>	<i>EVC</i>	<i>Total Variance %</i>
Student	44	0.258	0.87	0.547	1.39
Task	2	2.241	7.52	3.023	7.69
Rater	9	20.215	67.86	25.951	66.05
Student × Rater		1.429	4.80	2.255	5.74
Student × Task		0.207	0.69	0.317	0.81
Task × Rater		1.989	6.68	2.556	6.51
Student × Task × Rater		3.452	11.59	4.642	11.81
Generalizability Coefficient		0.26		0.57	

In Table 9, the universal score variance increased from 0.87% to 1.39%. This reflects slight differences between those two. The $s \times t$ interactions effect seems reduced from 67.86% to 66.05% and the $s \times r$ interaction seems increased from 4.80% to 5.74%. Slightly higher variance was obtained for differences in examinees' performance across tasks when the raters assigned scores by using GIM. Besides, the $s \times t$ interaction reduced from 6.68% to 6.51% when the raters assigned scores by using ESAS. However, a pretty higher generalizability coefficient was obtained when the scores were assigned using the scale. Moreover, the $s \times t \times r$ interaction increased from 11.59% to 11.81%. This might mean that inter-rater reliability is more effective and advantageous for revealing the differences in quality of students' responses when the scale is used to assign scores to the task.

Standardized Open-ended Questioning

Standardized open-ended questioning was employed for the instrumentation of the qualitative data in order to reveal the views of the raters on assessment processes and the types of assessments. It includes the same question –the same stimuli– in the same way determined in advance (Patton, 2002). The transcripts were analyzed line by line and memos were written (Strauss & Corbin, 1998; Glesne, 1999). Categories or labels were reviewed and recurring themes, core consistencies and meanings were

identified by using pattern codes (Miles & Huberman, 1994; Patton, 2002). The themes were found as : a) criteria use, b) spelling, and c) weightings

What is immediately apparent from open-ended transcripts is that the criteria use is very important and useful in essay assessment because the raters mention that they were more precise and the results were more consistent in assessing the essays by using the criteria given. One of the raters states that GIM assessments was like gambling because they needed to assign a total score to each essay without any written or pre-specified criteria. They also state that the criteria use changed the tendency of scoring subjectively in a positive manner. In this respect, raters seem to have the common idea those assessments by using a checklist or a scale is always more objective and reliable. Some teachers state that there should be a criterion for spelling. Even if the testees are advanced level writers, they might make spelling mistakes and the raters cannot score spelling because it is not one of the criteria in the scale. The spelling criterion had not been found appropriate by the judges because the task is at an advanced level. Although the raters seemed to have an agreement that GIM assessments were not reliable and consistent, they also criticized ESAS weightings. They state the criteria should not be equal for each sub-criterion. For example, one of the raters says it would be better if each weighting was different for each sub-criterion. In this way, it would be more useful and consistent. It would be particularly useful to state, considering the transcripts, that criteria use is a reliable and agreed measure for assessing essays. However, the criteria should be chosen precisely and correctly considering the needs of the students and the weightings of the criteria should be independent from each other. In fact, the weightings are different for each criterion but the particular teacher seems to think equal weightings are used for each criterion.

Discussion and Conclusions

The study gives evidence that all methods, techniques, or tools could include subjectivity and it seems reasonable to notice that mental processes and internal responses of raters function in different ways in using same assessment criteria for the same essays in different times. The statistical evidence indicates that GIM assessments are never consistent and reliable. The statistical analyses clearly show that ECC assessments are more reliable and consistent than GIM ones. The correlation coefficients are higher and they are supported by the raters themselves, as seen in qualitative data. The results also show that ESAS assessments are also consistent and reliable when compared to GIM. However, there is a slight difference between the correlation coefficients across ECC assessments and ESAS assessments. Yet, the coefficients across ESAS assessments are slightly higher and more meaningful than the ones across ECC assessments. This slight difference can also be observed by examining the intra-rater consensus between the assessments. It seems different weightings for each sub-criterion may result in more consistent assessments as raters declared because the results of the difference of correlation coefficients which were obtained by using Fischer's z transformation also support the idea that the intra-rater scores are similar but not the same. Paired comparisons with ANOVA

tell us the inter-rater scores are never meaningfully similar. This means different scores are assigned for the same essays in different time distances. It is obvious if a lower score is assigned to the same essay in two different sessions around the cut-off score, this means success and failure depend on a source of variation. At this point, the raters and the time elapsed between assessments may seem as the source of variation. The G coefficients also indicate that assigning scores is more precise and effective, when the scale is used, as it increases inter-rater reliability. Considering several limitations, further research into the effectiveness and usefulness of the scale would be valuable as it is difficult to infer what processes are experienced by the raters while they are scoring essays. The more pieces of information available, the more reliable will be the conclusions drawn from the data (Cherry & Meyer, 1993). However, when the total scores and the rater consensus results are examined, it can be clearly seen that the scores are different from each other even if the correlation coefficients are high and significant. It might be more accurate if Kline's (1986) cut-off coefficient (.70) for a meaningful correlation could be increased to .90 at least for giving evidence of more reliable ratings. This might mean the scores assigned are more similar and closer to each other. A deliberate training and agreement of raters before any process of rating for each student group also seems strongly needed on the criteria and performance indicators. In order to obtain verbal descriptions as concrete information, to recognize this process, and to establish the decision-making processes of raters, think-aloud protocols with follow-up interviews can also be employed.

References

- Blok, H. (1985). Estimating the reliability, validity, and invalidity of essay ratings. *Journal of Educational Measurement*, 22(1), 41-52.
- Bowen, K. and Cali, K. (2004). *Teaching the features of effective writing*. Retrieved November 21, 2004, from <http://www.learnnc.org/index.nsf/printView/1216418CB65B73CE85256D7300445C5A?OpenDocument>.
- Breland, H. (1983). The direct assessment of writing skill: A measurement review (Technical Report No.83-6). Princeton, NJ: College Entrance Examination Board.
- Celce-Murcia, M. (2001). *Teaching English as a second or foreign language*. Massachusetts: Heinle and Heinle.
- Chase, C. I. (1983). Essay test scores and reading difficulty. *Journal of Educational Measurement*, 20(3), 293-297.
- Chase, C. I. (1968). The impact of some obvious variables on essay test scores. *Journal of Educational Measurement*, 2(4), 315-318.
- Cherry, R. and Meyer, P. (1993). Reliability issues in holistic assessment. In M. Williamson and B. Huot (Ed.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 109-141). Cresskill, NJ: Hampton.

- Darus, S. (2006). Identifying dimensions and attributes of writing proficiency: development of a framework of a computer-based essay marking system for Malaysian ESL learners. *Internet Journal of e-Learning and Teaching*, 3(1), 1-25.
- Dempsey, M. S., PytlikZillig, L. M., and Bruning, R. G. (2009). Helping preservice teachers learn to assess writing: Practice and feedback in a web-based environment. *Assessing Writing*, 14, 38-61.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14, 88-115.
- Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Erkuş A. (2003). *Psikometri üzerine yazılar: ölçme ve psikometrinin tarihsel kökenleri, güvenilirlik, geçerlik, madde analizi, tutumlar; bileşenleri ve ölçülmesi* [Writings on Psychometrics: historical basis for measurement and psychometrics, reliability, validity, item analysis, attitudes; components and measurement]. 1. baskı, Ankara. Türk Psikologlar Derneği Yayınları.
- Fisher, R., Brooks, G., and Lewis, M. (2002). *Raising standards in literacy*. New York: Routledge.
- Glesne, C. (1999). *Becoming qualitative researchers: An introduction*. New York: Longman.
- Gyagenda, I. S. and Engelhard, G. (1998a). Rater, domain, and gender influences on the assessed quality of student writing using weighted and unweighted scoring. *Annual Meeting of the American Educational Research Association*. San Diego.
- Gyagenda, I. S. and Engelhard, G. (1998b). Applying the Rasch model to explore rater influences on the assessed quality of students' writing ability. *Annual Meeting of the American Educational Research Association*. San Diego.
- Hamp-Lyons, L. (1991). The writer's knowledge and our knowledge of the writer. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (p. 15-36). Norwood, NJ: Ablex.
- Hamp-Lyons, L. (1992). Holistic writing assessment for LEP students. *Second National Research Symposium on Limited English Proficient Student Studies: Focus on Evaluation and Measurement*. Washington.
- Hawkey, R. and Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing*, 9, 122-159.
- Herrington, A. and Moran, C. (2001). What happens when machines read our students' writing?. *College English*, 63, 480-499.
- Huang, J. (2012). Using generalizability theory to examine the accuracy and validity of large-scale ESL writing assessment. *Assessing Writing* 17, 123-139.

- Hughes, D. and Keeling, B. (1984). The use of model essays to reduce context effects in essay scoring. *Journal of Educational Measurement*, 21(3), 277-281.
- Hughes, D., Keeling, B., and Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement*, 20(1), 65-70.
- Hughes, D., Keeling, B., and Tuck, B. F. (1980). The influence of context position and scoring method on essay scoring. *Journal of Educational Measurement*, 17(2), 131-135.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60, 237-263.
- IELTS (2007). *IELTS handbook*. Retrieved January 19, 2008 from http://www.ielts.org/_lib/pdf/IELTS_Handbook_2007.pdf#.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., and Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Johnson, R., Penny, J., and Gordon, B. (2001). Score resolution and the interrater reliability of holistic scores in rating essays. *Written Communication*, 18(2), 229-249.
- Kan, A. (2005). Yazılı yoklamaların puanlanmasında puanlama cetveli ve yanıt anahtarı kullanımının (aynı) puanlayıcı güvenilirliğine etkisi [The effect of checklist and answer key use in writing assessment on rater reliability]. *Eğitim Araştırmaları Dergisi*, 5(20), 166-177.
- Kan, A. (2007). Performans değerlendirme sürecine katkıları açısından yeni program anlayışı içerisinde kullanılabilir bir değerlendirme yaklaşımı: Rubrik puanlama yönergeleri [An evaluation approach to be used for the new curriculum considering the contributions to performance evaluation process: Rubrics]. *Kuram ve Uygulamada Eğitim Bilimleri*, 7(1), 129-152.
- Kayapınar, U. (2010). A study on assessment tools and evaluation of essay writing skill in foreign language education. Unpublished PhD Dissertation, Mersin University. Yenisehir Campus: Turkey.
- Klein, S. and Hart, F. M. (1968). Chance and systematic factors affecting essay grades. *Journal of Educational Measurement*, 5(3), 197-206.
- Klein, J. and Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing Writing*, 10, 134-148.
- Kline, P. (1986). *A handbook of test construction: introduction to psychometric design*. London: Methuen.

- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26, 275-304.
- Marshall, J. C. and Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement*, 6(2), 97-101.
- Mertler, C. A. (2001). *Designing scoring rubrics for your classroom*. Practical Assessment, Research and Evaluation, 7(25). Retrieved October 11, 2007 from <http://PAREonline.net/getvn.asp?v=7&dn=25>.
- Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis*. California: Sage Publications.
- Moskal, B. M. (2000). *Scoring rubrics: what, when, and how?*. Practical Assessment, Research, And Evaluation, 7(3). Retrieved October 11, 2007 from <http://pareonline.net/getvn.asp?v=7&n=3>.
- Murphy, K. R. and Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology*, 74(4), 619-624.
- Nitko, A. J. (2001). *Educational assessment of students (3rd ed.)*. Upper Saddle River, NJ: Merrill.
- Norton, L. S. (1990). Essay-writing: What really counts. *Higher Education*, 20(4), 411-442.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. California: Sage Publications.
- Raimes, A. (1983). *Techniques in teaching writing*. Oxford: Oxford University Press.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.
- Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing* 17, 81-91.
- Strauss, A. and Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. California: Sage Publications.
- Sulsky, L. M. and Balzer W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology*, 73,497-506.
- Vaughan C. (1991). Holistic assessment: What goes on in the rater's mind? L. Hamp-Lyons (Ed.), *In Assessing Second Language Writing in Academic Contexts* (p. 111-126). Norwood, NJ: Ablex.
- Weir, C. J. (1990). *Communicative language testing*. Wiltshire: Prentice Hall.
- Wexley, K.N. and Youtz, M.A. (1985). Rater beliefs about others: Their effects on rating errors and rater accuracy. *Journal of Occupational Psychology*, 58, 265-275.
- Woehr, D. J. and Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.

Appendix 1:

ESSAY CRITERIA CHECKLIST (ECC)

-Make a checkmark if the essay includes the following attributes-

CRITERIA		CHECKMARK
ORGANIZATION	A. INTRODUCTION	
	A.1.1. Introductory Sentences	
	A.1.2. Thesis Statement	
	A.2. BODY PARAGRAPHS	
	A.2.1. Topic Sentence	
	A.2.2. Supporting Sentences	
	A.3. CONCLUSION	
LANGUAGE USE	B.1. Word Order	
	B.2. Pattern Variety	
	B.3. Verb Form	
	B.4. Tenses	
	B.5. Articles	
	B.6. Pronouns	
	B.7. Prepositions	
VOCABULARY	C.1. Word Choice	
	C.2. Word Variety	
	C.3. Parts of speech	
MECHANICS	D.1. Punctuation	
	D.2. Capitalization	
	D.3. Paragraphing	
	D.4. Indentation	
IDEAS/ CONTENT	E.1. Title	
	E.2. Development	
	E.3. Unity	
	E.4. Transitional Signals	

Appendix 2:

ESSAY ASSESSMENT SCALE (ESAS)

CRITERIA		ATTRIBUTES	4	3	2	1	0
ORGANIZATION	A.1. INTRODUCTION						
	A.1.1. Introductory Sentences	Effective introductory sentences					
	A.1.2. Thesis Statement	Appropriate thesis statement (thesis and central idea)					
	A.2. BODY PARAGRAPHS						
	A.2.1. Topic Sentence	Appropriate topic sentence (possibly implied) supporting the thesis and the central idea					
	A.2.2. Supporting Sentences	Appropriate sentences supporting the topic (possibly major and minor)					
	A.3. CONCLUSION	Appropriate conclusion related to thesis					
LANGUAGE USE	B.1. Word Order	Correct word order					
	B.2. Pattern Variety	Using different patterns					
	B.3. Verb Form	Using verb forms correctly					
	B.4. Tenses	Using tenses appropriately					
	B.5. Articles	Using articles correctly					
	B.6. Pronouns	Using pronouns correctly					
	B.7. Prepositions	Using prepositions correctly (verb + preposition, adjective + preposition)					
VOCABULARY	C.1. Word Choice	Selecting the appropriate words					
	C.2. Word Variety	Having a rich vocabulary					

	C.3. Parts of speech	Using the correct parts of speech					
MECHANICS	D.1. Punctuation	Using punctuation marks correctly					
	D.2. Capitalization	Using cases (lower/upper) correctly					
	D.3. Paragraphing	Correct paragraph formatting					
	D.4. Indentation	Using margins correctly and consistently					
IDEAS/ CONTENT	E.1. Title	Appropriate title					
	E.2. Development	Appropriate development					
	E.3. Unity	Unity					
	E.4. Transitional Signals	Using appropriate transitional signals					
TOTAL SCORE							

Kompozisyon Puanlamanın Ölçülmesi:

Aynı ve Farklı Puanlayıcı Güvenirliği

Atf:

Kayapınar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136
<http://dx.doi.org/10.14689/ejer.2014.57.2>

Özet

Problem Durumu: Yazma becerisinin etkili bir biçimde puanlanmasının araştırılmasına ilişkin bir hayli çaba gösterilmekte ve birçok öneri sunulmaktadır. Bu bağlamda, puanlayıcı güvenirligi, bireylerin gerek eğitim gerekse mesleki yaşamlarının farklı dönüm noktalarında hayati kararlar vermede çok önemli rol oynamaktadır. Aynı ve farklı puanlayıcıların farklı puanlama araçları kullanarak yaptıkları puanlamaların da güvenirlikleri puanlama süreçleri ile birlikte tartışılmalıdır.

Araştırmanın Amacı: Araştırmanın amacı İngilizce öğrencilerinin yazma becerilerinin aynı/farklı puanlayıcılar tarafından genel izlenim (GIM), kontrol listesi (ECC) ve kompozisyon puanlama ölçeği (ESAS) kullanılarak değerlendirilmesindeki olası farklılık ve tutarlılıkları ortaya çıkarmak ve puanlayıcı güvenirliklerini tartışmaktır.

Yöntem: Ölçme sonuçlarının tutarlılığı ve puanlamaların güvenirligine ilişkin yorum ve tartışmaların yapılabilmesi için nicel ve nitel veriler kullanılmıştır. Puanlama araçları 44 üniversite öğrencisi üzerinde uygulanmış ve 10 puanlayıcı genel izlenim, kontrol listesi ve ölçek kullanarak bu öğrencilerin yazma becerilerini puanlamışlardır.

Bulgular: Bulgular ve analiz sonuçları genel izlenimle puanlamanın beklendiği üzere kesinlikle güvenilir olmadığını göstermiştir. Elde edilen korelasyon katsayıları, varyans kestirimleri ve genellenebilirlik katsayılarından elde edilen bilgiler göz önüne alındığında, puanların aynı olmadığı ve sonuçlar arasında daima bir çeşitlilik ve varyasyon olduğu görülmektedir.

Sonuç ve Öneriler: Toplam puanlar ve puanlayıcıların vermiş oldukları puanlar arasındaki tutarlılıklar incelendiğinde sonuçların, korelasyon katsayıları yüksek ve anlamlı olsa dahi, çoğu zaman aynı olmadığı ve birbirlerinden farklı oldukları görülmüştür. Bu yüzden, yaygın kanının aksine, kontrol listeleri ve ölçekler, puanlayıcıların söz konusu araçlara yönelik iyi bir eğitim almamaları ve ölçütler, ölçüt tanımları ve performans göstergeleri üzerinde bir uzlaşma sağlamadıkları takdirde beklendiği gibi etkili bir şekilde güvenilir olamayabilmektedirler. Bu tür ölçmelerde anlamlı kabul edilecek korelasyon katsayısının en az .90 düzeyinde olması durumunda güvenilir puanlamaya kanıt oluşturacak olan sonuçlar daha hatasız olabilir. Bu durum aynı ve farklı yazılı yoklamalara verilen puanların birbirlerine olan yakınlık düzeylerini artıracak ve daha benzer sonuçların ortaya

çıkması anlamına gelebilecektir. Herşeye rağmen, hali hazırdaki durum ve sonuçlar gözönüne alındığında ölçek kullanımının diğer puanlama araçlarına göre daha güvenilir olduğu vurgulanabilir. Yine de çalışmanın daha fazla puanlayıcı ile tekrarlanmasının alana katkı sağlayacağı düşünülmektedir.

Anahtar Sözcükler: Kompozisyon, puanlama, puanlayıcılararası, puanlayıcı, güvenilirlik

Submission Checklist Makale Sunumu Kontrol Çizelgesi

Indicate that this submission is ready to be considered by this journal by checking off the following.

Aday makalenin değerlendirilmeye hazır olduğunu aşağıdakilerin her biri ile karşılaştırarak kontrol ediniz.

Manuscript Agreement: <i>I /We hereby accept that, the article after being accepted for publication in the Eurasian Journal of Educational Research (EJER), the author (s) as, all rights related to the article has been transferred to the ANI PUBLISHING under the laws of the , "copyright transfer".</i>	
Makale Eurasian Journal of Educational Research (EJER) dergisinde basıma kabul edildikten sonra, yazar(lar) olarak; makale ile ilgili tüm hakları, "Telif Hakkı Devir" yasaları uyarınca, ANI YAYINCILIĞA devrettiğimizi kabul ediyoruz.	
Manuscript Submission: <i>I have prepared my manuscript based on the criteria listed below and I accept all submission conditions.</i> Makalemi aşağıda belirtilen kriterlere göre hazırladım ve makale gönderme koşullarının tamamını kabul ediyorum.	
<i>Indicate that this submission is ready to be considered by this journal by checking off the following.</i> Aday makalenin değerlendirilmeye hazır olduğunu aşağıdakilerin her biri ile karşılaştırarak kontrol ediniz.	
1	<input checked="" type="checkbox"/> <i>The manuscript is a report of original educational research or a discussion article on research topics in education.</i> Aday makale, eğitim alanıyla doğrudan ilgili bir araştırma ya da tartışma yazısıdır.
2	<input checked="" type="checkbox"/> <i>The submission has not been previously published, nor sent to another journal for consideration.</i> Sunulan çalışma daha önce herhangi bir yerde yayınlanmamıştır, başka bir derginin incelemesinde değildir.
3	<input checked="" type="checkbox"/> <i>Within a year, I have not submitted a manuscript to EJER as an author or co-author for review other than this manuscript.</i> Son bir yıl içerisinde, yazarı olduğum ya da yazarları arasında bulunduğum başka bir çalışma değerlendirilmek üzere EJER'e sunulmamıştır.
4	<input checked="" type="checkbox"/> <i>All authors of this manuscript are subscribers of the EJER and accept to pay 300 EURO for the cost of proofreading and 10 loose copies of their articles if the manuscript is accepted.</i> Bu aday makalenin tüm yazarları EJER'e abonedirler ve aday makalenin kabul edilip basılması halinde makalelerinin İngilizce son okuma (proofreading) ve 10 serbest kopya gönderimi nedeniyle oluşan 675 TL maliyeti Anı Yayıncılık'a ödemeyi kabul ederler.

5	<input checked="" type="checkbox"/>	<i>The entire manuscript is written in English.</i> Aday makalenin bütününün yazım dili İngilizce'dir.
6	<input checked="" type="checkbox"/>	<i>The original manuscript is typed on A4 paper. The margins are 2.5 cm.</i> Aday makale kenar boşlukları 2.5 cm olan A4 kağıda yazılmıştır.
7	<input checked="" type="checkbox"/>	<i>Each paragraph is longer than two sentences.</i> Her bir paragraf en az üç cümle içermektedir.
8	<input checked="" type="checkbox"/>	<i>The entire manuscript - including quotations, references, author note, content footnotes, figure captions, and all parts of tables - is double-spaced.</i> Aday makalenin tamamı, alıntılar, kaynakça, şekil ve tablo başlıkları da dahil olmak üzere çift aralıklı yazılmıştır.
9	<input checked="" type="checkbox"/>	<i>The submission file is in Microsoft Word document file format. 12-point Times New Roman font is used in entire manuscript.</i> Aday makale, tamamında 12 punto Times New Roman yazı tipi kullanılarak hazırlanmış bir Microsoft Word dokümanıdır.
10	<input checked="" type="checkbox"/>	<i>The text has had the authors' names removed. If an author is cited, "Author" and year are used in the bibliography and footnotes, instead of author's name, paper title, etc. The author's name has also been removed from the attached document.</i> Aday makale, yazar adları çıkarılarak sunulmuştur. Eğer yazar kendisine atıfta bulunduysa yazarın adına ve çalışma başlığına yer verilmeyecek, sadece "Author" yazılarak çalışmanın yılı belirtilecektir. Eklenen dosyada yazar adı belirtmeyecektir.
11	<input checked="" type="checkbox"/>	<i>The title is 10 to 12 words.</i> Aday makalenin başlığı 10-12 sözcük uzunluğundadır.
12	<input checked="" type="checkbox"/>	<i>The maximum length of the manuscript-including structured abstract in English, tables, and references is 6000 words. This limitation does not include Turkish extended abstract (750-1000 words) which is placed after the references section.</i> Aday makale, İngilizce abstract, tablolar ve kaynakça vb. tüm öğeler dahil olmak üzere en fazla 6000 sözcüktür. Kaynakça'nın ardından yer verilen uzun Türkçe özet (750-1000 sözcük) bu sayıya dahil değildir.
13	<input checked="" type="checkbox"/>	<i>The article is preceded by English Structured Abstract of not more than 400 words and not less than 300 using five required headings: Problem Statement, Purpose of Study, Methods, Findings and Results, and Conclusions and Recommendations. (These headings may need some adaptation in the case of discussion papers: Background, Purpose of Study, Sources of Evidence, Main Argument, and Conclusions). More information available from (http://www.tandf.co.uk/journals/authors/rereabstracts.asp)</i> Yapılandırılmış İngilizce öz 300-400 sözcük uzunluğunda olup, aday makalenin başında yer almakta ve <i>Problem Durumu, Araştırmanın Amacı, Araştırmanın Yöntemi, Araştırmanın Bulguları, Araştırmanın Sonuçları ve Önerileri</i> başlıklarını içermektedir. Bu başlıklar tartışma yazıları için: <i>Çalışmanın Temeli, Çalışmanın Amacı, Kanıt Kaynakları, Ana Tartışma ve Sonuçlar</i> şeklinde olabilir. Daha fazla bilgi için http://www.tandf.co.uk/journals/authors/rereabstracts.asp adresine başvurunuz.

14	<input checked="" type="checkbox"/>	<p>Following the structured abstract in English four to six keywords are included.</p> <p>Yapılandırılmış İngilizce özden sonra 4-6 anahtar sözcüğe yer verilmiştir.</p>
15	<input checked="" type="checkbox"/>	<p>An extended (750-1000 words) Turkish structured abstract is placed following the "References" section using five required headings: Problem Statement, Purpose of Study, Methods, Findings and Results, and Conclusions and Recommendations. (These headings may need some adaptation in the case of discussion papers: Background, Purpose of Study, Sources of Evidence, Main Argument, and Conclusions). More information available from http://www.tandf.co.uk/journals/authors/rereabstracts.asp</p> <p>Kaynakça'dan sonra 750-1000 sözcükten oluşan Türkçe yapılandırılmış öze yer verilmiştir. Türkçe yapılandırılmış öz <i>Problem Durumu, Araştırmanın Amacı, Araştırmanın Yöntemi, Araştırmanın Bulguları, Araştırmanın Sonuçları ve Önerileri</i> başlıklarını içermektedir. Bu başlıklar tartışma yazıları için: <i>Çalışmanın Temeli, Çalışmanın Amacı, Kanıt Kaynakları, Ana Tartışma ve Sonuçlar</i> şeklinde olabilir. Daha fazla bilgi için; http://www.tandf.co.uk/journals/authors/rereabstracts.asp</p>
16	<input checked="" type="checkbox"/>	<p>Following the Turkish structured abstract, four to six keywords are included.</p> <p>Uzun Türkçe özetten sonra 4-6 anahtar sözcüğe yer verilmelidir.</p>
17	<input checked="" type="checkbox"/>	<p>References are not cited in the structured abstracts in English and in Turkish.</p> <p>İngilizce abstract ve Türkçe öz içerisinde atıfta bulunulmamıştır.</p>
18	<input checked="" type="checkbox"/>	<p>The format of headings, tables, figures, citations, references, and other details follow the APA 6 style as described in the <i>Publication Manual of the American Psychological Association, 6th edition</i>, available from http://www.apa.org</p> <p>Aday makalenin başlıkları, tabloları, şekilleri, atıfları, kaynakçası ve diğer özellikleri tamamen APA altıncı baskıda belirtildiği şekildedir.</p>
19	<input checked="" type="checkbox"/>	<p>All illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.</p> <p>Aday makalenin şekilleri ve tabloları metin içerisinde bulunmaları gereken uygun yerlere yerleştirilmiştir. Makale sonunda sunulmamıştır.</p>
20	<input checked="" type="checkbox"/>	<p>Citations in the text of the document include the author's surname, the year of publication, and, when there is a specific quote from a source used, a page number where the quote is located in the text.</p> <p>Example: Nothing seemed so certain as the results of the early studies (Tatt, 2001, p. 445). It was precisely this level of apparent certainty, however, which led to a number of subsequent challenges to the techniques used to process the data (Jones & Wayne, 2002, p. 879). There were a number of fairly obvious flaws in the data: consistencies and regularities that seemed most irregular, upon close scrutiny (Aarns, 2003; West, 2003, p. 457).</p> <p>With studies by two authors, always include both author names:</p>

		<p>(Anderson & Bjorn, 2003) As Anderson and Bjorn (2003) illustrated in their recent study As recently as 2003, a prominent study (Anderson & Bjorn) illustrated When a study has 3, 4, or 5 authors, include the names of all the authors the first time the work is cited: (Anderson, Myers, Wilkes, & Matthews, 2003) For all subsequent citations of this work, use "et al." (Anderson et al., 2003) When a work has 6 or more authors, use et al.: (Bell et al., 2003) For unsigned works, include the title, enclosed in parentheses. Put quotation marks for short work titles, and italicize the titles of reports, books, and other significant works: ("Recent Developments," 2004) (Dictionary of Tetrathalocigistic Diseases, 2004)</p> <p>Metin içindeki atfları üstte verilen örneklere uygundur.</p>
21	<input checked="" type="checkbox"/>	<p>Three levels of headings are used: Level 1, Level 3 and Level 4. The headings are formatted as follows: Centered Uppercase and Lowercase Heading (Level 1) <i>Flush Left, Italicized, Uppercase and Lowercase Side Heading</i> (Level 3) <i>Indented, italicized, lowercase paragraph heading ending with a period.</i> Start writing after the period (Level 4).</p> <p>Aday makale içerisinde üç farklı düzey başlık kullanılmıştır. Düzey 1, Düzey 2, Düzey 3. Başlıklar bu düzeylere uygun olarak aşağıdaki şekilde biçimlendirilmiştir:</p> <p>Ortalı ve Her Sözcüğün İlk Harfi Büyük Yazılmış Başlık (Düzey 1) <i>Tam Sola Dayalı, İtalik ve Her Sözcüğün İlk Harfi Büyük Yazılmış Başlık</i> (Düzey 3) <i>İçeriden, itatik, tamamı küçük harflerle yazılmış ve nokta ile bitten başlık.</i> Noktadan sonra normal metin yazımına devam edilmeli (Düzey 4).</p>
22	<input checked="" type="checkbox"/>	<p>References are listed in alphabetical order. Each listed reference is cited in text, and each text citation is listed in the References. Basic formats are as follows: Haag, L., & Stern, E. (2003). In search of the benefits of learning Latin. <i>Journal of Educational Psychology</i>, 95, 174-178. Bollen, K. A. (1989). <i>Structural equations with latent variables</i>. New York: Wiley. Johnson, D. W., & Johnson, R. T. (1990). Cooperative learning and achievement. In S. Sharan (Ed.), <i>Cooperative learning: Theory and research</i> (pp. 173-202). New York: Praeger.</p> <p>Turkish References Only: Çınkır, Ş., & Çetin, S. K. (2010). Öğretmenlerin okullarda mesleki çalışma ilişkileri hakkındaki görüşleri [Teachers' opinions about the professional working relationships in schools]. <i>Kuram ve Uygulamada Eğitim Yönetimi</i>, 16(3), 353-371.</p>

	<p>Article in an Internet-only journal/Periodical, database Fredrickson, B. L. (2000, March 7). Cultivating positive emotions to optimize health and well being. <i>Prevention & Treatment</i>, 3, Article 0001a. Retrieved November 20, 2000, from http://journals.apa.org/prevention/volume3/pre0030001a.html</p> <p>More information is available from: http://citationonline.net/CitationHelp/csg04-manuscripts-apa.htm#references</p> <p>Kaynakçanın yazımı üstte verilen örneklere uygundur.</p>
23	<p><input checked="" type="checkbox"/> <i>Order of the main parts in the manuscript is as follows:</i> <i>Main title in English (max. 12 words)</i> <i>Structured abstract (min. 300- max.400 words length)</i> <i>Keywords (in English, min. four-max. six)</i> <i>Main text</i> <i>References</i> <i>Main Title in Turkish (max. 12 words)</i> <i>Extended structured abstract (min.750-max.1000 words length in Turkish)</i> <i>Keywords (in Turkish, min. four-max. six)</i></p> <p>Aday makaleyi oluşturan ana öğeler aşağıdaki düzendedir: İngilizce Ana Başlık (En fazla 12 sözcük) Yapılandırılmış İngilizce Abstract (En az 300, en fazla 400 sözcük) Anahtar Sözcükler (İngilizce, en az dört, en fazla altı) Ana Metin Kaynakça Türkçe Ana Başlık (En fazla 12 sözcük) Yapılandırılmış Türkçe Öz (En az 750, en fazla 1000 sözcük) Anahtar Sözcükler (Türkçe, en az dört, en fazla altı)</p>

24	☒	<p>Structure of the Manuscript should be as follows:</p> <p>Manuscript Title</p> <p>English Abstract (350-400 words) Sub-headings:</p> <p><i>Problem statement:...</i></p> <p><i>Purpose of Study:...</i></p> <p><i>Method:...</i></p> <p><i>Findings and Results:...</i></p> <p><i>Conclusions and Recommendations:...</i></p> <p><i>Keywords:..</i></p> <p>Introduction Method (sub-headings):</p> <p><i>Research Design</i></p> <p><i>Research Sample</i></p> <p>♦<i>Research Instrument and Procedure</i></p> <p><i>Data Analysis</i></p> <p>Results, Discussion and Conclusion, References</p> <p>Extended Turkish Abstract (750-1000 words) Sub-headings:</p> <p><i>Problem Durumu:....</i></p> <p><i>Araştırmanın Amacı:....</i></p> <p><i>Araştırmanın Yöntemi:...</i></p> <p><i>Araştırmanın Bulguları:...</i></p> <p><i>Araştırmanın Sonuçları ve Önerileri:.....</i></p>
25	☒	<p><i>EJER editor reserves the right to reject manuscripts and not to publish it that do not comply fully with the EJER Manuscript Submission Guidelines.</i></p>

ABONELİK

Yıllık

Kişi Aboneliği : 80 TL

Kurum Aboneliği : 250 TL

Yurtdışı Aboneliği: : 250 \$

Öğrenci Aboneliği: : 70 TL

Tek Sayılar İçin:

Kişiler : 25 TL

Kurumlar : 75 TL

ÖDEME

İş Bankası Kızılay Şubesi, Anı Yayıncılık 4214-0962561
numaralı hesap

♦ Reliability and the validity of the research instrument used or adapted in the work must be provided, and explained in detail.