METACOGNITIVE CALIBRATION IN COLLEGE STUDENTS: THE RELATIONSHIP BETWEEN CALIBRATION, TEST PERFORMANCE AND EMPATHIC SKILLS

ESRA ERDEM

BOĞAZİÇİ UNIVERSITY

METACOGNITIVE CALIBRATION IN COLLEGE STUDENTS: THE RELATIONSHIP BETWEEN CALIBRATION, TEST PERFORMANCE AND EMPATHIC SKILLS

Thesis submitted to the

Institute for Graduate Studies in the Social Sciences

in partial satisfaction of the requirements for the degree of

Master of Arts

in

Psychology

by

Esra Erdem

Boğaziçi University

2012

Thesis Abstract

Esra Erdem, "Metacognitive Calibration in College Students: The Relationship Between Calibration, Test Performance, and Empathic Skills"

Metacognition has been argued to play an important role in learning and performance in academic settings. Most of the self-regulated learning theories consider metacognitive calibration ability as central to successful learning. Calibration can be defined as one's ability to evaluate one's performance and is regarded as one of the most critical elements differentiating between capable and less capable learners. Recently, it has been argued that for metacognitive calibration to be more accurate, individuals should take the external criteria as reference in their evaluation of their performance. In this thesis, the main question of interest was whether and how metacognitive calibration contributed to students' performance in a school setting. In addition, students' calibration ability against an external criterion (in this case, the instructor) has been assessed in terms of their empathic ability. Results showed that better academic performance was positively correlated with higher metacognitive accuracy. In addition, better performing students showed underconfidence and poorer students showed overconfidence in their metacognitive judgments. There was no relationship between emphatic skills of the students and their metacognitive ability.

Tez Özeti

Esra Erdem, "Üniversite Öğrencilerinde Üstbilişsel Kalibrasyon: Kalibrasyon, Test Performansı ve Empati Becerisi Arasındaki İlişki"

Üstbilişin, eğitimle ilgili alanlarda öğrenme ve performans üzerine önemli rol oynadığı tartışılmaktadır. Özdüzenleyici öğrenme teorilerinin çoğu, üstbilişsel kalibrasyon yeteneğini başarılı bir öğrenmenin merkezinde görürler. Kalibrasyon, bir kişinin kendi performansını değerlendirebilme yeteneği olarak tanımlanabilir ve yüksek öğrenme kabiliyetine sahip olanlar ve olmayanları birbirinden ayıran kritik unsurlardan biri olarak görülür. Şu sıralar, daha kesin üstbilişsel kalibrasyon için bireylerin kendi performans değerlendirmelerine referans olabilecek harici ölçütler kullandıkları tartışılmaktadır. Bu tezde, üstbilişsel kalibrasyonun okul ortamındaki öğrencilerin performanslarını nasıl etkilediği sorusu ile ilgilenilmiştir. Buna ek olarak, harici bir ölçütün (bu durumda, öğretmen) var olduğu durumlarda öğrencinin kalibrasyon yeteneği empati yeteneği üzerinden ölçülmüştür. Sonuçlar göstermiştir ki, daha iyi bir akademik performans ile üstbilişsel yargılarının kesinliği ilişkisi pozitif yönde anlamlıdır. Ek olarak, akademik performansı daha iyi olan öğrencilerin üstbilişsel yargılarında az kendine güven (underconfidence) gösterdikleri ve performansı daha kötü olan öğrencilerin üstbilişsel yargılarında fazla kendine güven (overconfidence) gösterdikleri bulunmuştur. Öğrencilerin empati becerisi ve üstbilişsel yetenekleri arasında anlamlı bir ilişki bulunmamıştır.

CONTENTS

CHAPTER I

INTRODUCTION

The aim of the study is to determine the relation between metacognitive calibration and performance in a school context. Additionally, this study aims to investigate the potential link between calibration and perspective taking/empathy skills.

One of the earliest definitions of metacognition refers to the experiences and knowledge we have about our own cognitive processes (Flavell, 1979). It involves "knowledge about the nature of people as cognizers, about the nature of different cognitive tasks, about possible strategies that can be applied to the solution of different tasks and it also involves executive skills for monitoring and regulating one's cognitive activities" (Flavell, 1999, p.21). In line with Flavell's definitions, Kluwe (1982; as cited in Hacker, 1998) and Palincsar and Brown (1987) also acknowledge two components in the definition of metacognition: monitoring and control.

Regarding the relation between monitoring and control, Nelson and Narens (1990, 1994; as cited in Nelson, 1996) formulated a theoretical mechanism consisting of two structures: meta-level and object-level. Object-level is defined by the existence of external objects and meta-level refers to our cognitions related with these external objects. In this system, meta-level is informed by the object level and this is called as "monitoring". Correspondingly, object-level is modified by the meta-level through a continuing communication flow and this is called as "control". Some examples of monitoring component are ease of learning judgments, judgments of learning, feeling of knowing judgments, and confidence in retrieved answers and

some examples of control component are selection of kind of processing, allocation of study time, termination of study, and termination of search (Nelson & Narens, 1994). In this paper, our focus will be on monitoring component of metacognition.

Self-Regulated Learning and Calibration

In terms of the role of metacognition in education, it is suggested that metacognition fosters effective control of learning (Metcalfe, 2009); if students know the current state of their knowledge and make correct judgments of what is known and not known, they can effectively manage their learning. Indeed, there is substantial evidence that proficient learners are those who employ diverse metacognitive processes over the cognitive entity (Allen & Armour-Thomas, 1993). Metacognitive ability has also been shown as an important individual difference between capable and less capable learners (Baker, 1989; Brown & Campione, 1986; Garner & Alexander, 1989; Pressley & Ghatala, 1990; as cited in Everson & Tobias, 1998).

Metacognitive theory contributes to students' understanding of themselves as being active agents of their own learning (Hacker, 1998). This point is critical, since it directly taps on the concept of "self-regulated learning (SRL)". Self-regulation implies the individual's autonomy and control in monitoring, directing, and regulating his/her actions during the course of a cognitive task (Paris & Paris, 2001). The individual's decisions to handle the task depends on his/her metacognitive knowledge of task features and task demands as well as his/her metacognitive experiences, feelings, and judgments related to the task (Efklides, 2009).

Calibration is one aspect of metacognitive monitoring in the sense that it is an evaluation of an individual's perception of his/her own performance and it is a component of self-regulated learning (Pieschl, 2009). Congruence between students' predictions of their capabilities in a learning task (metacognitive judgments) and actual performances on that task tells us how much calibrated an individual is concerning the task (e.g., Glenberg & Epstein, 1985, 1987; Lin & Zabrucky, 1998; Schraw & Roedel, 1994; Stone, 2000; Weingardt, Leonesio, & Loftus, 1994; Garavalia & Gredler, 2003; as cited in Pieschl, 2009). Regarding educational considerations, it has been shown that there is a relation between metacognitive judgments and achievement levels of students, and the literature provides data about metacognitive calibration skills of high versus poor performing students. Generally high performing students are found to be more calibrated compared to their low performing counterparts and there is a tendency for high performing students to display underconfidence, whereas low performing students usually display overconfidence (e.g., Kruger & Dunning, 1999; Bol & Hacker, 2001; Hacker, Bol, Horgan, & Rakow, 2000). So, calibration accuracy is considered as an important individual difference, which is related to self-regulation and a predictor of school achievement (Winne & Jamieson-Noel, 2002).

Calibration judgments are typically realized by obtaining confidence judgments or achievement estimates. These judgments have generally been collected in two ways: local or global. Whereas local level refers to the confidence judgment of a single item, global level refers to the prediction of overall performance on the task. As an example, Nietfeld, Cao, and Osborne's (2005) study measured

monitoring accuracy of their subjects at both local and global level across three multiple question tests and a final exam which consisted items in various difficulty levels in an educational psychology course. Students gave confidence ratings for each test item as they answered each question by drawing a slash through the portion of a 100-mm line that best matched to their perceived confidence; one end of the line corresponded to no confidence 0% accurate and the other end corresponded to total confidence/100% accurate. This was the local monitoring measurement. For the global monitoring measurement, students did the same task considering their overall performance on the test. There are many other types of calibration research, but typically, global measurements are obtained by asking students to estimate how many questions they believed they answered correctly immediately after they finished the test and for local measurements, confidence judgments were assessed for each item.

Calibration is generally measured by accuracy and bias indexes. Calibration accuracy is calculated by taking the absolute value of the difference between confidence judgments and performance for each item and calibration bias is calculated by taking the signed difference between the average confidence and average performance score, which in turn yields underconfidence/overconfidence (Nietfeld et al., 2005). Calibration judgments can be made both before (called as prediction) and after (called as postdiction) taking a task. Actually, postdictions are proved to be more accurate compared to predictions in almost all cases (Maki, 1998).

Considering the relation between calibration accuracy and performance, it is said that exposing the same experience through similar testing sessions increases performance on various metacognitive tasks (Thiede, 1999; Nelson, Dunlosky, Graf, & Narens, 1994; Metcalfe, 2002, Metcalfe & Kornell, 2003, 2005; Son & Metcalfe, 2000; Butler & Roediger, 2007; Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006; as cited in Kelemen, Winningham, & Weaver, III, 2007). Supporting this argument, in their study, Kelemen et al. (2007) asked their subjects to make JOLs over 5 experimental sessions which included the study of list of different words and they found a significant improvement in metacognitive performance over the 5 consecutive sessions. However, there is a counter argument which suggests that practice does not lead to an increase in metacognitive performance (Bol & Hacker, 2001; Gigerenzer, Hoffrage, & Kleinbolting, 1991; Koriat, 1997; Koriat, Lichtenstein, & Fischoff, 1980; as cited in Bol, Hacker, O'Shea, & Allen, 2005). For examle, in their study, Bol et al. (2005) assigned their subjects to two different groups: practice and no-practice groups. In the practice group, subjects were asked to make predictions and postdictions regarding their performance over 5 exams throughout a semester, whereas no-practice group were not required to do so. Actually it turned out that practice effect did not work and it did not resulted in a siginificant improvement in metacognitive performance. Therefore, one of the aims of the current study will be to reveal the practice effect over multiple exam sessions in a real classroom setting.

Metacognitive Adaptation to External Criteria

Relation between metacognitive calibration and school performance have been generally measured through the above mentioned traditional measures of calibration; predictions and postdictions at the item levels. However one of the important points of this study is the fact that apart from these traditional measures of metacognitive judgments, a new measure, which has been introduced in the calibration research recently, will also be used in order to reveal the relation between calibration and performance - the degree of "calibration to external criteria" (Pieschl, 2009).

Knowing what one knows and does not know about a text/item is a measure of awareness of one's own cognition, but for successful self-regulation, one should be able to monitor external demands and make necessary cognitive and metacognitive regulations according to the relevant external criteria. In line with this, it is suggested that "thinking about other people's thinking" should be included as another metacognitive measure to increase the individual's own metacognitive awareness (Cooper & Stewart, 2006).

In educational settings, relevant external criteria is generally assumed to be the teachers' perception of task and learning material (e.g., Broekkamp, van Hout-Wolters, van den Bergh, & Rijlaardam, 2004; Jetton & Alexander, 1997; Hadwin, 2008). Students should take teachers' thinking processes related to the instructional stimuli into account while regulating own cognitive processes. This point directs us toward a fundamental issue to discuss in educational contexts; task understanding. Effective self-regulation is the primary driver of school performance, but meeting at the same understanding of any task/instructional material by the teacher and the

student is the prerequisite for effective self-regulation and learning process. Especially, students are exposed to task/test demands which are generally complex and not simply answered at the college/university level. Therefore, student calibration should be conceptualized not only by means of confidence/bias judgments, but also degree of match between the student and teacher regarding the understanding of the task should be included in calibration studies.

Along with task understanding, improvement of calibration in classroom contexts should be related to a particular kind of task demand; determining importance in school material. This competency has a metacognitive dimension, because the individual has to engage in higher order cognitive processes and reflect upon the to be learned material in order to decide which parts of the text are important and which are unimportant. Actually, reflection and analysis of thought, drawing conclusions from this analysis, and applying these conclusions into the practice are central to the concept of metacognition (Downing, Kwong, Chan, Lam, & Downing, 2009). Supporting this view, the relation between the selection of important units of a written material and metacognitive skills was investigated in a study by Brown and Smiley (1977), in which children of four different age group (8, 10, 12, and 18) were asked to rate the important elements of a text. Results showed that when compared to older children, younger ones were less successful in the task and this failure was attributed to immature metacomprehension skills of young children. However, the study did not examine the relation between the selection of important units of a written material and school performance.

In instructional settings, students are generally expected to assess what their teachers consider important in texts, since teachers are the immediate context for task demands and test questions (Hadwin, Oshige, Miller, & Wild, 2008). In one of the studies, experimenters asked a group of secondary school students and teachers to select the main points of a biology text by underlying the important and most important text elements. Investigators were mainly interested in the degree of correspondence between teacher and student perceptions of importance in the text (Schellings & van Hout-Wolters, 1995). Single words and word groups were seperately analyzed. Although the results revealed that mean difference of students' and teachers' selections of main points was not significant, this was due to large standart deviations. Actually, it turned out that there was a large variation between teacher and student perceptions of importance. Many other studies have been conducted to examine the teacher and student perceptions of "importance" in educational psychology literature, but the general conclusion in this body of research supports the view that students have difficulty in selecting main points that will be selected by their teachers.

A more recent study comes from Broekkamp, van Hout-Wolters, van den Bergh, and Rijlaardam (2002). They examined students' and teachers' selection of important points from a history book chapter contained approximately 8000 words (80 pages), but found little correspondence between the teacher and student ratings. Researchers emphasized the need for students to be well-attuned to task demands set by their teacher in order to succeed in their assignments. Interestingly, as stressed in Broekkamp et al.'s (2002) study, there is not much evidence of the relation between

students' perceptions of task demands (task demands are assumed to hold for "teacher perception") and school performance in the literature. Therefore, this question constitutes the other important investigation of this research paper. We will try to find out whether good-calibration with the teacher perception will lead to better performance. However, the task representing "external calibration measure" in the current study is different from those mentioned so far, because data will be collected from a group of students enrolled in a statistical lesson, so it is not reasonable to ask the students to select/underline the "important parts" of a text as statistics is more a quantitative/mathematical lesson, rather than a content lesson. Actually, the difference of this study here refers to the complexity of the statistical material, which does not have that clear-cut an evaluation, compared to the evaluation of the match of the list of words in a learning task.

Linking Calibration with Social Cognition

As implicitly emphasized throughout the above studies, metacognition has a social aspect, because it involves monitoring of the other person's cognition as well as own cognition. Actually, this aspect of metacognition forms the basis of the social underpinnings and functioning of "theory of mind" research (Efklides, 2009). We know that representation of mental states of others and perspective taking ability are central for the definition of theory of mind concept. In this sense, it is very tight to empathy, since empathy can be viewed as the same with "perspective taking" (Preston & de Waal, 2002).

Empathy has both affective and cognitive dimensions (Kerem, Fishman, & Josselson, 2001). Whereas the affective side of empathy involves shared emotions, experiences, and vicarious feeling, cognitive side involves perspective or role-taking. Cognitive component is also suggested to be synonymous with using a *theory of mind* (Baron-Cohen, 1995; Wellman, 1990; as cited in Billington, Baron-Cohen, & Wheelwright, 2007). It is obvious that a student's understanding of task demands or selection of main points is only possible if the student monitors the teacher's mental states during learning process. Furthermore, traditional calibration measurements, such as indicating the degree of confidence in one's answers in an exam or self-evaluation of one's own answers by giving achievement estimates, can not be dissociated from its social cognitive aspects, because a student must take into account the teacher's perspective as he/she is the evaluator of the student's performance. Therefore, calibration and empathic (perspective taking) skills should also be related in the school context.

Another social cognitive phenomenon, perceived similarity, has also been investigated in this study as one of the factors that may be related with metacognitive calibration accuracy and the actual performance of students. As suggested by many social psychologists, inferring mental states is very central to social life and it drives much of causal explanation (e.g. Kruglanski, 1975; Malle, 2001; Sutton & McClure, 2001; as cited in Ames, 2004). In this case, perceived similarity with the teacher has been studied as a factor relating to metacognitive calibration accuracy and the actual performance of students.

Study 1:

In the light of the research reviewed, the study 1 first attempts to examine the relation between calibration and test performance through the two types of metacognitive tasks in a real classroom environment (local/global postdictions and adaptation to teacher perspective). Then, the link between metacognitive calibration and empathy (perspective taking) skills is questioned. Although some researchers (e.g. Pieschl, 2009) referred to the importance of social cognition for effective self-regulation and calibration, to our knowledge there is no published empirical work on the topic. In this sense, the study will be unique in highlighting the dynamic relationship among calibration, performance and empathy (perspective taking) in university students.

Study 2:

The study 2 first examines the relation between the level of development of metacognitive skills and test performance. Secondly, as in the study 1, metacognitive calibration and test performance has been investigated in a real classroom environment. However, in this study, local/global postdictions have been analyzed during the full semester, so that we could investigate the practice effect. Then similar to *the study 1*, the link between metacognitive calibration and empathy (perspective taking) skills is questioned. Finally, the relation between perceived similarity and calibration accuracy as well as test performance have also been examined.

The Hypotheses

Based on the previous ideas of the current paper, it can be hypothesized that students who have more developed metacognitive skills are expected to have greater metacognitive calibration accuracy and perform better in the exam. In the same vein, students with better calibration are expected to be more successful in the exam. Moreover, low performing students are expected to display overconfidence, whereas high performing ones are expected to display underconfidence. It is also expected that practice effect will lead to increased test performance.

Secondly, well-calibrated students (those who have greater metacognitive calibraovtion accuracy) are expected to have more developed empathic (perspective taking) skills. Similarly, it is expected that students' perceived similarity with the teacher is correlated with metacognitive calibration accuracy and test performance.

CHAPTER II

METHOD

Study 1

Participants

Participants were 38 undergraduate students enrolled in "Introduction to statistics for psychology II" lesson. Students got extra course credit for their participation in the study.

Measures and Procedure

Confidence or Achievement Judgments.

Immediately after finishing the final exam, students were asked to indicate to rate their overall performance on the exam, which yielded global monitoring accuracy (postdiction). Then, they were asked to give a confidence judgment for each question in terms of the degree of their expectation of each question's probability to be included in the exam. They provided confidence judgments between 0-100 as a measure of external calibration measure. Followingly, students evaluated their performance against each question's score (item-specific) that yielded local monitoring accuracy (postdiction). Difference between the actual realization of performance and estimation of performance provided the basis of local judgments and global monitoring accuracy. Finally, students completed *The Interpersonal Reactivity Index* an empathy measure.

The Interpersonal Reactivity Index.

This questionnaire (IRI) was developed by Davis (1980). According to Davis' model, empathy has both affective and cognitive dimensions and consists of a set of distinct, but related constructs. Cognitive constructs are fantasy (the tendency to project one's self into the feelings and actions of fictitious characters portrayed in movies, books, etc.) and perspective taking (Davis, 1980). Affective constructs are empathic concern (other-oriented feelings of sympathy and concern) and personal distress (self-oriented feelings of anxiety and unease (Davis, 1980).

The IRI is a self-report instrument and it has 28 items. As stated above, it has four subscales: fantasy, perspective taking, empathic concern, and personal distress. Items are rated on a 4-point Likert-type scale (0=does not describe me well 4=describes me very well). The IRI was adapted into Turkish and this version was used in the current study (Engeler & Yargıç, 2007). For Turkish adaptation of the instrument, data was collected from a total of 214 undergraduate students. Internal reliability is .76 for phantasy subscale, .66 for empathic concern subscale, .73 for perspective taking subscale, and .60 for personal distress subscale. Test-retest reliabilities are between .66 and .80 ($p<.0001$). The full version of the questionnaire is in Appendix A.

Study 2

Participants

Participants were 56 undergraduate students enrolled in “Introduction to statistics for psychology II” lesson. Students got extra course credit for their participation in the study.

Measures and Procedure

Study 2 was almost the same with the study 1; however, we excluded the external calibration measure – students’ prediction of each question’s probability to be included in the exam, since it did not provide sufficient varience. Instead of this measure, after finishing with the global and local monitoring judgments, students completed the Turkish adaptation of Metacognitive Awareness Inventory (MAI), which was developed by Schraw & Dennison’s (1994). It is named as *Yetişkinler için Bilişüstü Testi.*

Yetişkinler için Bilişüstü Beceri Testi

This questionnaire (YBBT) was developed by (Ozcan, 2007). It is a two-factor model measuring metacognitive awareness: Knowledge of Cognition and Regulation of Cognition. It is composed of 52 questions and internal reliability measure Cronbach's alpha was reported as .95. Items are rated on a 5-point Likert-type scale (0=Never 5=Allways). The total score was taken as a measure of the level of metacognitive development. Test-retest reliability was .998 ($p<.01$).

Perceived Similarity Items.

To measure the perceived similarity with the target (the teacher), students responded to 2-questions, which were included in YBBT questionnaire. The questions were as follows: “I think I’m very similar to the instructor of this course,” “The instructor of this course is a lot different from me” [reversed]; participants indicated their agreement with these items on a 6-point scale ranging from 1 [very strongly disagree] to 6 [very strongly agree]). The full version of the questionnaire is in Appendix B.

CHAPTER III

RESULTS

Study 1

The first question that was investigated is the Study 1 is the relation between local and global monitoring judgments. Local judgments are calculated by the following way: the sum of all the differences between the actual and predicted scores at the item level are divided by the total number of exam items. Global judgments are calculated by the difference between the actual and predicted scores considering the total exam score. There was a significant correlation between local and global judgments, $r = .908$, $p(\text{two-tailed}) < .01$, which indicated that the students were highly consistent with their judgments at the local and global levels.

Calibration and Performance Relation

Considering the relation between global monitoring judgment accuracy and performance, it has been found that performance scores were negatively correlated with the absolute differences between predicted and actual scores, $r = -.435$, $p(\text{two-tailed}) < .01$. This indicated that high performance is positively correlated with greater accuracy (smaller difference).

Considering the relation between local monitoring judgment accuracy and performance, it has been found that performance scores were negatively correlated with the absolute differences between predicted and actual scores, $r = -.426$, $p(\text{two-tailed}) < .01$. This again indicated that high performance is positively correlated with greater accuracy smaller difference).

In order to examine over/underconfidence of subjects, signed differences of actual and predicted scores were also used as a measure of metacognitive accuracy, which is named as judgment bias. It provides the information about the direction of the difference. Figure 1 in Appendix C shows that high performing students displayed underconfidence, whereas low performing students displayed overconfidence.

Calibration with the External Criteria and Performance Relation

To reveal the relation between calibration with the external criteria and performance, students were asked to give a confidence judgment for each question in terms of the degree of their expectation of each question's probability to be included in the exam ranging from 0 to 100. However, at the specific item (question) level, it turned out that there was no significant correlation between calibration with the external criteria and performance, $p > .05$. The only significant relation was between the average of overall confidence judgments and performance, $r = .388$, $p(\text{two-tailed}) < .05$. Table 1 in Appendix D shows the correlation matrix of calibration with the external criteria (confidence ratings) and performance relation.

Calibration Accuracy and Empathy (Perspective Taking) Skills Relation

To reveal the relation between calibration accuracy and empathy (perspective taking) skills, students were asked to fill out the IRI questionnaire. There were four subscales of the questionnaire: fantasy, perspective taking, empathic concern, and

personal distress. However, it turned out that there was no significant correlation between these subscales and calibration accuracy, $p > .05$.

Table 2 in Appendix E shows the correlation matrix of calibration accuracy and empathy (perspective taking) skills relation.

Study 2

A summary of the descriptive statistics for the study variables is presented in Table 3 in Appendix F. The correlation matrix for all study variables is presented in Table 4 in Appendix G.

The first question that was investigated in the Study 2 is the relation between local and global monitoring judgments. Local and global monitoring judgments are calculated as the same way in Study 1. There was a significant correlation between local and global monitoring judgments for the 1st and 2nd midterms and final exam, $r = .85$, $p(\text{two-tailed}) < .01$, $r = .92$, $p(\text{two-tailed}) < .01$, $r = .77$, $p(\text{two-tailed}) < .01$, respectively. This again indicated that the students were highly consistent with their judgments at the local and global levels.

Relation between the Level of Development of Metacognitive Skills and Test Performance

Student were asked to fill out YBBT questionnaire to examine the relation between the level of development of metacognitive skills and test performance, which yielded no significant correlation for the 1st midterm and final exam, $p > .05$. However, for

the 2nd midterm, a significant correlation has been found, r =.30, p(two-tailed) < .05.

Calibration and Performance Relation

Considering the relation between calibration and performance for the 1st midterm and final exam, we have found no significant correlation at both local and global level, p> .05. However, for the 2nd midterm, we have found that performance scores were negatively correlated with the absolute differences between predicted and actual scores at the local level, r = -.43, p(two-tailed) < .01 and global level, r = -.33, p(two-tailed) < .05. This again indicated that high performance is positively correlated with greater accuracy (smaller difference).

Practice Effect

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that showed that there was a significant effect of time on calibration accuracy, F (1.725, 110) = 4,198, $p< .05$. Posthoc comparisons revealed that there was a significant difference in calibration accuracy between the 1st midterm *($M= .44$, $SD= .06$)* and final exam *($M= .74$, $SD= .09$)*, $p< .01$. Since calibration accuracy was calculated as the difference between predicted and actual scores, higher mean score in the final exam indicated a decrease in calibration accuracy when compared to the 1st midterm. Other comparisons did not demonstrate significant differences.

Over/Underconfidence and Performance Relation

Subjects were divided into 2-groups based on the difference between their expected and actual scores for the 1st, 2nd midterm and the final exam. If the difference was (-), they were classified as underconfident group and if the difference was (+), they were classified as overconfident group (+). Perfect accuracy (=0) was not included into the analyses.

For the first midterm, an independent samples subjects t-test showed that the actual performance of the underconfident group (*M= 18,97, SD= 1,61)* was significantly higher than that of the overconfident group (*M= 14,17, SD= 2,69), t(39)= 7,128, p<* .001).

For the second midterm, an independent samples subjects t-test showed that the actual performance of the underconfident group (*M= 16,50, SD= 1,79)* was significantly higher than that of the overconfident group (*M= 13,00, SD= 2,65), t(43)= 4,793, p<* .001).

For the final exam, an independent samples subjects t-test showed that the actual performance of the underconfident group (*M= 17,60, SD= 4,08)* was significantly higher than that of the overconfident group (*M= 14,77, SD= 3,06), t(47)= 2,654, p<* .001).

Calibration Accuracy and Empathy (Perspective Taking) Skills Relation

Table 4 in Appendix G shows the correlation matrix of calibration accuracy and empathy (perspective taking) skills relation. Among the all correlation between the four subscales of IRI questionnaire and calibration accuracy for the three exams, the

only significant correlation were between phantasy scale and calibration accuracy (measured as the difference between the predicted and actual scores at the local lovel) for the 1st midterm and final exam, $r = .32$, $p(\text{two-tailed}) < .05$ and $r = .30$, $p(\text{two-tailed}) < .05$, respectively. However, this positive correlation did not support our hypothesis considering the relation between empathic skills and calibration accuracy.

Calibration Accuracy, Test Performance and Perceived Similarity Relation

Table 4 in Appendix G shows the correlation matrix of calibration accuracy and perceived similarity relation. For the 1st midterm, it turned out that perceived similarity was only positively correlated with the actual test performance, $r = .36$, $p(\text{two-tailed}) < .05$ and for the 2nd midterm, perceived similarity was only positively correlated with the actual test performance, $r = .35$, $p(\text{two-tailed}) < .05$. Other correlations did not demonstrate a significant relation.

CHAPTER IV

DISCUSSION

The purpose of this study was to determine the relation between metacognitive calibration and performance in a school context. It also aimed at exploring the potential link between calibration and perspective taking/empathy skills. In the first study, students' performances were analyzed over the final exam and in the second study, performances were analyzed over the two midterms and the final exam.

Study 1

The results of the study 1 suggested that there was a strong link between calibration accuracy and test performance. This result is consistent with many of the literature findings regarding metacognitive judgment and performance relation (e.g., Nietfeld et al., 2005; Hacker et al., 2000; Horgan, 1990; Fitzgerald, Gruppen, White & Davis, 1997; as cited in Grimes, 2002). It is clear that students' metacognitive judgment ability is an important differentiating factor in terms of performance level. The results also supported our hypothesis in the sense that underconfident students performed better when compared to overconfident students. This is again consistent with the literature which examines over/underconfidence debate in various contexts (e.g., Hacker et al., 2000; Kruger & Dunning, 1999; Bol et al., 2001).

Previously, it was stated that the teachers are the relevant external criteria at school context (e.g., Broekkamp, et al., 2004; Jetton, et al., 1997; Hadwin, 2008). Therefore, it was predicted that good-calibration with the teacher perception would lead to better performance. In order to investigate this relation, we asked the students

to give a confidence judgment between 0-100 (percentage) for each question in terms of the degree of their expectation of each question's probability to be included in the exam. Contrary to our assumption, we could not find a relation between calibration with the external criteria and performance at the local (question) level. The only significant relation we had was at the global level (the average of the total confidence judgments). Actually, this difference at local and global levels may be deriving from the fact that at the question level, there was really very limited variation among the confidence judgments, so it did not help to reveal any potential relation. Similarly, we could not detect a significant relation between the subscales of IRI questionnaire and calibration accuracy. Theoretically, it seems that calibration accuracy is not related with empathic skills or we may not be measuring what we really intended to measure with this questionnaire: perspective taking ability. The questionnaire has not been replicated yet and it may need further validation. Actually, it would be better to explore this relation with an experimental task which specifically used for measuring adult perspective taking skills. One of the other potential explanations of this finding is, as in the case of calibration analyses of external criteria discussed above, the range of scores used in calibration accuracy was quite limited because of the low variation of scores of exam questions (between 1-6). Again, this might be one of the factors which prevented us to reveal any potential relation with empathic (perspective taking) skills.

Study 2

In the study 2, we first examined the relation between the level of development of metacognitive skills, which was measured by the YBBT questionnaire, and test performance. It turned out that among the three exams, we had significant correlation between the level of development of metacognitive skills and test performance for only one exam. For the rest of the other exams, we could not find a significant relation. To say that these two variables are not related seems counterintuitive. However, one explanation may be linked to the fact that YBBT is a relatively newly developed questionnaire (Ozcan, 2007) and it has not been replicated in other studies yet, as in the case of IRI questionnaire. Therefore, it may need to be validated and improved with further studies. Indeed, there is some support for the contention, at least for one of those exams.

Similarly, considering the relation between calibration accuracy and performance, we had significant correlation at both local and global level for only one exam, which was supported by the fact that high performing students had higher calibration accuracy. For the rest of the other exams, we could not find a significant relation. This is really unexpected, since this is a very established phenomenon in metacognition literature. However, for both of the three exams, we could identify that higher performers showed more underconfidence than poorer performers, which was remarked by their underestimation of test performances.

Another research question that was examined in this study was the practice effect. We examined practice effect over the three exams throughout the full semester. The litarature on practice effect has mixed results. Some of the researchers

support the view that there is a practice effect which is evident by the increasing metacognitive calibration accuracy and test performance over time and with repeated exposure to same experience (e.g., Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006; as cited in Kelemen, et al., 2007). However, our findings supported the counter view, since we could not find an increasing calibration accuracy or test performance over time and with repeated testing sessions. What is more interesting is the fact that it turned out that test performance decreased significantly over time and with repeated testing sessions. Although this is quite surprising, we can see similar results when we look at the literature. For example, Bol et al. (2001) grouped students in introductory research methods in education course into two groups: practice and no-practice group. They found that students who took practice tests before the final exam, such as taking midterms before the final exam as in our case, performed worse than the group which did not take practice test. One of the explanation of this finding that the researchers provided was the fact that by taking practice tests, students had an expectation that the practice test would be nearly identical with the actual exam, but when they met with different questions, both their metacognitive calibration accuracy and test performance decreased as a result of the differency between the two tests. Another study, which was conducted among a group of university student enrolled in an education course, supports this finding. In this study, students again were assigned to practice and no practice groups. The practice group took five quizzes before the final exam. However, they found that there was no significant difference between the two groups in terms of metacognitive calibration accuracy

and test performances. Our finding and these studies can be suggested as another evidence for the ongoing debate which supports the view that calibration accuracy is resistant to improvement (Bol et al., 2001; Dembo & Jakubowski, 2003; Gigerenzer et al., 1991; Koriat, 1997; as cited in Bol et al., 2005).

Regarding the relation between empathic skills and metacognitive calibration accuracy, again we could not find a significant relation between the two variables.We may propose the same arguments in Study 1 as an explanation to this finding. However, as this is the second application of the same questionnaire on a different population, we can accept that there is no relation between empathic skills and calibration accuracy. Similarly, we could not find a significant relation between perceived similarity with the target (teacher) and metacognitive calibration accuracy. The only significant relation was between perceived similarity and the actual test performance. We need further data to validate these results and come to a conclusion in terms of the relation between perceived similarity, calibration and performance.

To summarize, our results support that there is a strong relation between test performance and calibration accuracy. Moreover, high performing students display underconfidence, whereas low performing students display overconfidence. These findings are compatible with the metacognition literature. However, one of our main interests in this research was to explore a potential link with social aspect of metacognition and calibration, because this would be the first study showing the relation between calibration, test performance and social cognition. Unfortunately, the results did not support our hypothesis. Apart from the explanations mentioned above, one of the reasons that prevented us to reveal such a relation may be deriving

from the fact that we collected data from a statistical course, however if it were a content lesson, we could found a different pattern, because in content lessons, there are open ended questions and there is more room for discussion of different views. Therefore, we could have the opportunity to see that good calibration with the teacher perspective or perceived similarity with the teacher were related with calibration accuracy. It would be interesting to make further research with a larger population and with a population from a content lesson.

APPENDICES

APPENDIX A

Kişilerarası Tepkisellik İndeksi

Aşağıdaki ifadeler sizin değişik durumlardaki düşüncelerinizi ve duygularınızı soruşturmaktadır. Her bir maddenin sizi ne kadar iyi tanımladığını gösteren sayıyı, belirtilen puanlamayı referans alarak işaretleyiniz.

1 2 3 4 5
Beni iyi bir şekilde tanımlıyor — Beni çok tanımlamıyor

1. Başıma gelebilecek olan şeyler hakkında, zaman zaman hayaller ve fanteziler kurarım.	1	2	3	4	5
2. Benden daha talihsiz insanlar için genellikle merhametli, alakalı hisler duyarım.	1	2	3	4	5
3. Olayları "bir başka kişinin" bakış açısından görmeyi zor bulurum.	1	2	3	4	5
4. Başka kimselerin problemleri olduğunda, onlar için fazla üzülmem.	1	2	3	4	5
5. Bir romandaki karakterlerin duygularını gerçekten içimde hissederim.	1	2	3	4	5
6. Acil durumlarda, vesveseli ve rahatsız hissederim.	1	2	3	4	5
7. Bir piyes veya film izlerken genellikle tarafsızımdır ve sıklıkla kendimi ona tamamen kaptırmam.	1	2	3	4	5
8. Bir karara varmadan önce diğerlerinin anlaşamadığı yönlerden olaya bakmaya çalışırım.	1	2	3	4	5
9. Birinden yararlanıldığını gördüğümde, ona karşı koruyucu olduğumu hissederim.	1	2	3	4	5
10. Çok heyecanlı bir durumun içinde olduğumda çaresizlik hissederim.	1	2	3	4	5
11. Arkadaşlarımın bakış açısından olayların nasıl göründüğünü gözümde canlandırarak onları daha iyi anlamaya gayret ederim.	1	2	3	4	5
12. İyi bir kitaba veya filme son derece kapılmak benim için bir parça nadir bir durumdur.	1	2	3	4	5
13. Birinin incindiğini gördüğümde, sakin kalma eğilimindeyimdir.	1	2	3	4	5
14. Başka kimselerin talihsizlikleri genellikle beni büyük ölçüde rahatsız etmez.	1	2	3	4	5
15. Bir şeyde haklı olduğumdan eminsem, başkalarının fikirlerini dinleyerek fazla zaman harcamam.	1	2	3	4	5
16. Bir piyes veya filmi gördükten sonra, karakterlerden biriymişim gibi hissetmişimdir.	1	2	3	4	5
17. Gergin duyguların olduğu bir ortamda olmak beni	1	2	3	4	5

korkutur.	
18. Birine haksız davranıldığını gördüğümde, onlar için bazen çok fazla acıma hissetmem.	1 2 3 4 5
19. Genellikle acil durumların üstesinden gelmede çok becerikliyimdir.	1 2 3 4 5
20. Gördüğüm şeyler bana oldukça dokunur.	1 2 3 4 5
21. Her sorunun iki yönü olduğuna inanırım ve her iki yönden de bakmaya çalışırım.	1 2 3 4 5
22. Kendimi oldukça yumuşak kalpli bir kişi olarak tanımlarım.	1 2 3 4 5
23. İyi bir film seyrettiğimde, kendimi çok kolaylıkla baş karakterin yerine koyabilirim.	1 2 3 4 5
24. Acil durumlarda kontrolü kaybetmeye eğilimliyimdir.	1 2 3 4 5
25. Birine kızdığımda, genellikle bir süre için kendimi onun yerine koymaya çalışırım.	1 2 3 4 5
26. İlginç bir hikaye veya roman okuduğumda, hikayedeki olaylar benim başıma gelse neler hissedeceğimi gözümde canlandırırım.	1 2 3 4 5
27. Acil bir durumda çok yardıma ihtiyacı olan birini gördüğümde paramparça olurum.	1 2 3 4 5
28. Birilerini eleştirmeden önce, onların yerinde olsam nasıl hissedeceğimi gözümün önünde canlandırmaya çalışırım.	1 2 3 4 5

APPENDIX B

Yetişkinler İçin Biliş Üstü Beceri Testi

Aşağıdaki maddeler yetişkinlerin öğrenirken biliş üstü becerileri kullanıp kullanmadıklarını anlamaya yöneliktir. Dikkatlice okuyup her maddenin sizin için ne kadar geçerli olduğunu (X) işareti koyarak belirtiniz.

		Asla	Nadiren	Bazen	Sık Sık	Her Zaman
1	Kendime düzenli aralıklarla amaçlarıma ulaşıp ulaşmadığımı sorarım.					
2	Bir probleme cevap vermeden önce çeşitli alternatifleri dikkate alırım					
3	Geçmişte işe yaramış olan stratejileri					
4	Öğrenirken yeterli zamana sahip olabilmek					
5	Zihinsel olarak zayıf ve güçlü yönlerimi					
6	Bir işe başlamadan önce gerçekten ne öğrenmeye ihtiyacım olduğunu düşünürüm.					
7	Bir sınavı bitirdiğimde ne kadar iyi yaptığımı					
8	Bir işe başlamadan önce özel amaçlarımı					
9	Önemli bir bilgiyle karşılaştığımda					
10	Öğrenmek için hangi tür bilginin en önemli					
11	Problem çözerken kendime bütün seçenekleri dikkate alıp almadığımı sorarım.					
12	Bilgiyi organize etmede iyiyim.					
13	Bilinçli olarak dikkatimi önemli olan bilgiye					
14	Kullandığım her stratejide kesin bir amacım					
15	Bir konuyu en iyi, o konu hakkında bilgi sahibi olduğumda öğrenirim.					

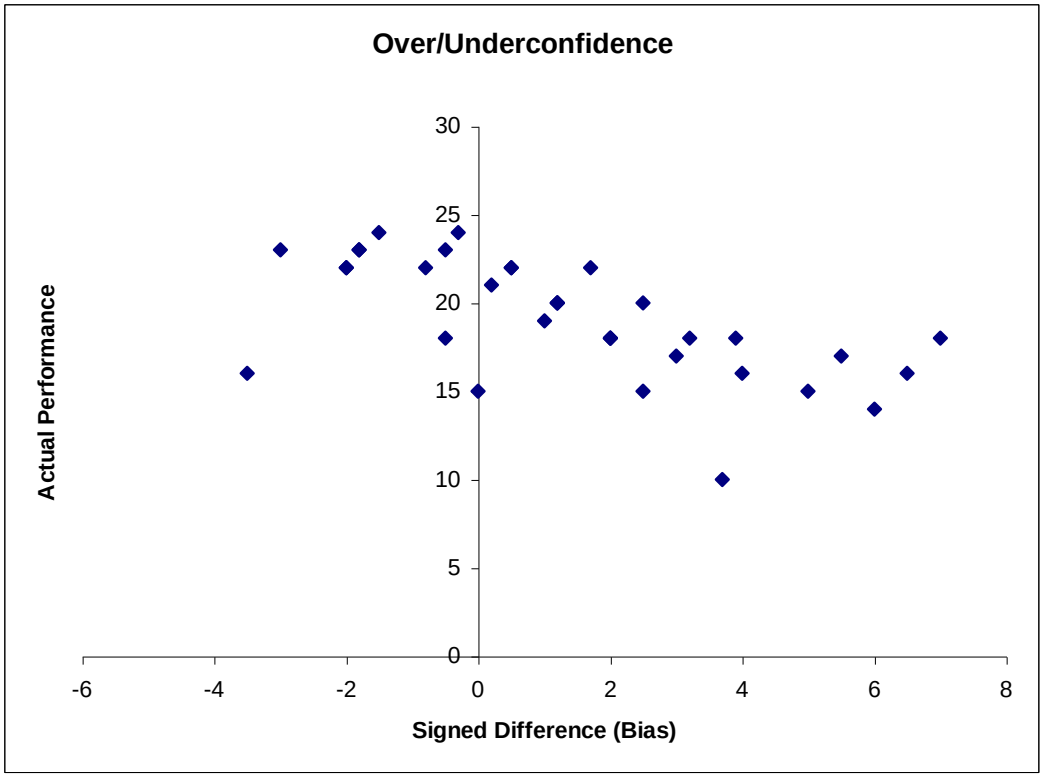
16	Bir şeyi öğrenmem için benden ne					
17	Bilgiyi hatırlama konusunda iyiyimdir.					
18	Koşullara uygun olarak değişik öğrenme					
19	Bir işi bitirdikten sonra bu işi yapmak için daha kolay bir yol olup olmadığını kendime sorarım.					
20	Ne kadar iyi öğrendiğimi kontrol edebilirim.					
21	Önemli bağlantıları anlamama yardımcı olması için düzenli aralıklarla tekrar yaparım.					
22	Bir konuya başlamadan önce o konu ile ilgili kendi kendime sorular sorarım.					
23	Problem çözerken en iyisini seçebilmek için birçok yol düşünürüm.					
24	Öğrendiklerimi bitirdikten sonra özetlerim.					
25	Bir şeyi anlamadığımda başkalarından yardım alırım.					
26	Bir şeyi öğrenmeye ihtiyacım olduğunda bunun için kendimi motive edebilirim.					
27	Çalışırken hangi stratejileri kullandığımın farkındayım					
28	Herhangi bir konuyu çalışırken kullandığım stratejilerin yararlı olup olmadığını analiz ederim.					
29	Zayıf yönlerimi kapatmak için zihinsel olarak güçlü yönlerimi kullanırım.					
30	Yeni bir konuyu öğrenirken, o bilginin anlamına ve önemine yoğunlaşırım.					
31	Bir bilgiyi daha anlamlı kılmak için kendi örneklerimi yaratırım.					
32	Herhangi bir şeyi ne kadar iyi anladığımı					

	değerlendirebilirim.					
33	Yararlı öğrenme stratejilerini kendiliğinden kullanırım.					
34	Bir konuyu anlayıp anlamadığımı düzenli ararlıklarla durup kontrol ederim.					
35	Kullandığım her bir stratejinin ne zaman en etkili olacağını bilirim.					
36	Bir işi bitirdiğimde amaçlarıma ne kadar ulaştığımı kendime sorarım.					
37	Öğrenirken anlamama yardımcı olması için resimler ve şekiller çizerim.					
38	Bir problemi çözdükten sonra bütün seçenekleri dikkate alıp almadığımı kendime sorarım.					
39	Yeni bir bilgiyi kendi cümlelerimle aktarırım.					
40	Anlamakta zorlandığımda stratejilerimi değiştirim.					
41	Öğrenmeme yardımı olması için, metnin yapısını düzenlerim.					
42	Bir işe başlamadan önce o işle ilgili açıklamaları dikkatlice okurum.					
43	Okuduğumun daha önceki bildiklerimle bağlantısı olup olmadığını kendime sorarım.					
44	Aklım karıştığında varsayımlarımı tekrar değerlendiririm.					
45	Amaçlarımı en iyi şekilde başarmak için zamanımı organize ederim.					
46	Konuya ilgi duyduğumda daha iyi öğrenirim.					
47	Çalışmayı daha küçük basamaklara bölmeye çalışırım.					

48	Bir konuyu ayrıntılarından ziyade genel hatlarıyla ele alırım.					
49	Yeni bir şeyi öğrenirken ne kadar iyi yaptığımı kendi kendime sorarım.					
50	Bir işi bitirdiğimde yeteri kadar öğrenip öğrenemediğimi kendime sorarım.					
51	Anlaşılmayan bir bilgiyi derinlemesine araştırırım.					
52	Anlamadığımda durur ve tekrar okurum.					

APPENDIX C

Figure 1. Over/underconfidence as displayed by the relation between actual performances and signed differences



APPENDIX D

Table 1

The correlation matrix of calibration with the external criteria (confidence ratings) and performance relation.

	Actual Performance
Conf.R._1st_Q	.146
Conf.R._2nd_Q	.302
Conf.R._3rd_Q	.093
Conf.R._4th_Q	.177
Conf.R._5th_Q	.086
Overall _Conf.R.	.388*

Conf.R._1st_Q : Confidence rating for the 1st question

Conf.R._2nd_Q : Confidence rating for the 2nd question

Conf.R._3rd_Q : Confidence rating for the 3rd question

Conf.R._4th_Q : Confidence rating for the 4th question

Conf.R._5th_Q : Confidence rating for the 5th question

Overall _Conf.R. : Average confidence rating for the total exam

* Correlation is significant at the 0.05 level (2-tailed)

APPENDIX E

Table 2

The correlation matrix of calibration accuracy and empathy (perspective taking) skills relation.

	Absolute Difference
Phantasy Scale	.088
Perspective Taking Scale	.097
Empathy Scale	.218
Distress Scale	.288

Absolute Difference: Measure of calibration accuracy

APPENDIX F

Table 3. Descriptive statistics for the variables in Study 2

	N	M	SD	Min.	Max.
FirstLocalSD	56	-0.02	0.64	-1.42	2.67
FirstLocalAD	56	0.44	0.46	0	2.67
FirstGlobalSD	56	-0.47	3.90	-10	16
FirstGlobalAD	56	2.56	2.97	0	16
FirstActual	53	16.74	3.38	8	20
SecondLocalSD	56	0.24	0.73	-1.1	3.4
SecondLocalAD	56	0.51	0.57	0	3.4
SecondGlobalSD	56	0.97	3.66	-5.5	16
SecondGlobalAD	56	2.56	2.77	0	16
SecondActual	52	13.96	3.14	7.5	19
FinalLocalSD	56	0.41	0.91	-2	3.6
FinalLocalAD	56	0.74	0.67	0	3.6
FinalGlobalSD	56	2.19	4.78	-8	18
FinalGlobalAD	56	3.99	3.40	0	18
FinalActual	52	15.31	3.75	6.5	23
Phantasy	52	26.42	4.80	11	35
Perspective	52	24.94	4.21	14	34
Empathy	52	26.79	3.42	11	35
Distress	52	22.31	5.07	12	32
MetacognitiveSkill	51	179.04	21.59	117	231
SocialProjection	47	3.45	1.35	1	6
FirstLocalEstimation	56	1.38	0.49	1	2
Estimation2	49	1.43	0.50	1	2
LocalAD_averaged	56	0.56	0.33	0.1	1.49

APPENDIX G

Table 4. Correlation matrix for the variables in Study 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1)1.LocalAD		-,24	,87**	,12	,09	-,04	,13	-,05	-,03	,32*	,19	,19	,22	,03	-,05
2)1.Actual			-,15	,05	,06	,47**	-,03	,06	,55**	,03	-,26	,00	-,06	,22	,36*
3)1.GlobalAD				,07	,10	,01	,04	-,06	-,02	,17	,16	,22	,20	,02	-,00
4)2.LocalAD					,93**	-,43**	-,17	-,24	,17	,18	-,16	,09	-,14	-,07	,14
5)2.GlobalAD						-,33*	-,16	-,19	,24	,24	-,17	,13	-,19	-,15	17
6)2.Actual							,20	,33*	,24	,01	-,11	,20	,25	,30*	,35*
7)3.LocalAD								,77**	-,20	,30*	,08	-,00	-,10	-,15	-,15
8)3.GlobalAD									-,12	,18	,04	-,10	-,03	,03	-,01
9)3.Actual										,05	-,26	,10	-,13	,14	,07
10)Phantasy											,11	,28*	,20	-,10	,02
11)Perspective Taking												,19	-,06	,21	,15
12)Empathy													,24	,10	,24
13)Distress														-,07	,03
14)YBBT															,27
15)Perceived Similarity															

1)1.LocalAD: First Midterm Local Level Absolute Differences
2)1.Actual: First Midterm Actual Total Exam Score
3)1.GlobalAD: First Midterm Global Level Absolute Differences
4)2.LocalAD: Second Midterm Local Level Absolute Differences
5)2.GlobalAD: Second Midterm Global Level Absolute Differences
6)2.Actual: Second Midterm Actual Total Exam Score
7)3.LocalAD: Final Exam Local Level Absolute Differences
8)3.GlobalAD: Final Exam Global Level Absolute Differences

9)3.Actual: Final Exam Actual Total Exam Score
10-13)Phantasy, Perspective Taking, Empathy, Distress: Subscales of IRI questionnaire
14)YBBT: Turkish adaptation of Metacognitive Awareness Inventory (MAI)
15)Perceived Similarity: Perceived similarity with the target

---

* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

REFERENCES

Allen, B. A., & Armour-Thomas, E. (1993). Construct validation of metacognition. *The Journal of Psychology, 127 (2),* 203-211.

Ames, D. R., (2004). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology, 87 (5),* 573-585.

Billington, J., Baron-Cohen, S., & Wheelwright, S. (2007). Cognitive style predicts entry into physical sciences and humanities: Questionnaire and performance tests of empathy and systemizing. *Learning and Individual Differences, 17,* 260-268.

Bol, L. & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69 (2),* 133-151.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73 (4),* 269-290.

Broekkamp, H., van Hout-Wolters, B. H. A. M., van den Bergh, H., & Rijlaardam, G. (2002). Importance in instructional text: Teachers' and students' perceptions of task demands. *Journal of Educational Psychology, 94 (2),* 260-271.

Broekkamp, H., van Hout-Wolters, B. H. A. M., van den Bergh, H., & Rijlaardam, G. (2004). Teachers' task demands, students' test expectations, and actual test content. *British Journal of Educational Psychology, 74,* 205-220.

Brown, A. L. & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development, 48,* 1-8.

Cooper, S. S. & Stewart, P. W. (2006). Metacognitive development in professional educators. Paper presented at the annual convention of the American Educational Research Association in San Francisco.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10,* 85.

Downing, K., Kwong, T., Chan, S.-W., Lam, T.-F., & Downing, W.-K. (2009). Problem-based learning and the development of metacognition. *Higher Education, 57 (5),* 609-621.

Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema, 21 (1),* 76-82.

Efklides, A. (2009). The new look in metacognition: From individual to social, from cognitive to affective. In C. B. Larson (Ed.), *Metacognition: New Research Developments* (pp 137-151). New York: Nova Science Publishers.

Engeler, A. & Yargıç, İ. L. (2007). Kişilerarası tepkisellik indeksi: Empatinin çok boyutlu ölçümü, *New Symposium Journal, 45 (3),* 119-27.

Everson, H. T & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26,* 65-79.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34 (10),* 906-911.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50,* 21-45.

Grimes, (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *The Journal of Economic Education, 33 (1),* 15-30.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp 1-25). Mahwah, NJ: Lawrence-Erlbaum.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92 (1),* 160-170.

Hadwin, A. F. (2008). Do your students really understand your assignments? *Tomorrow's Professor Blog.* Massachusetts Institute of Technology and Stanford University. http://ampstools.mit.edu/tomprofblog/archives/2008/03/857_do_your_stu.html

Hadwin, A. F., Oshige, M., Miller, M., & Wild, P. (2008). Examining student and instructor task perceptions in a complex engineering design task. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), New York.

Kelemen, W. L., Winningham, R. G., & Weaver, III, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology, 19 (4/5),* 689-717.

Kerem, E., Fishman, N., & Josselson, R. (2001). The experience of empathy in everyday relationships: Cognitive and affective elements. *Journal of Personal and Social Relationships, 18 (5),* 709-729.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in

recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77 (6),* 1121-1134.

Maki, R. H. (1998). Text predictions over text material. In D. J. Hacker, J. Dunlosky, A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp 117-144). Mahwah, NJ: Lawrence-Erlbaum.

Metcalfe, J. (2009). Metacognitive judgements and control of study. *Current Directions in Psychological Science, 18 (3),* 159–163.

Nelson, T. O. & Narens, L. (1994). Why investigate metacognition. In J. Metcalfe & A.P. Shimamura (Eds.), *Metacognition, knowing about knowing,* (pp 1-25). Cambridge: The MIT Press.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51 (2),* 102-116.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. The *Journal of Experimental Education, 74 (1),* 7-28.

Ozcan, Z. C., (2007). Sınıf öğretmenlerinin derslerinde bilişüstü beceri geliştiren stratejileri kullanma özelliklerinin incelenmesi. *Doctoral dissertation,*Marmara University, Istanbul, Turkey.

Palincsar, A. S., & Brown, D. A. (1987). Enhancing instructional time through attention to metacognition. *Journal of Learning Disabilities, 20 (2),* 66-75.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36 (2),* 89-101.

Pieschl, S. (2009). Metacognitive calibration – an extended conceptualization and potential applications. *Metacognition Learning, 4,* 3-31.

Preston, S. D. & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25 (1),* 1-71.

Schellings, G. L. M. & van Hout-Wolters, B. H. A. M (1995). Main points in an instructional text, as identified by students and their teachers. *Reading Research Quarterly, 30 (4),* 742-756.

Winne, P. H. & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27,* 551-572.