The Effectiveness of Source Monitoring Training in Reducing Halo Effect and Negativity Bias in a Performance Appraisal Setting

by

Hatice İrem Gökcel

Koç University

A Thesis Submitted to the
Graduate School of Social Sciences
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts

in

Psychology
Koç University
December 2007

Koç University

Graduate School of Social Sciences

This is to certify that I have examined this copy of a master's thesis by

Hatice İrem Gökcel

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____
Tarcan Kumkale, Ph. D. (Advisor)

_____
H. Canan Sümer, Ph. D.

_____
Mahmut Bayazit, Ph. D.

_____
Institute of Social Sciences, Director

Date: _____

STATEMENT OF AUTHORSHIP

This thesis contains no material which has been accepted for any award or any other degree or diploma in any university or other institution. It is affirmed by the candidate that, to the best of her knowledge, the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signed Hatice İrem Gökcel

ABSTRACT

Performance appraisals are regularly conducted in both work organizations and educational settings. However, rater biases in performance appraisal are quite common, and decrease the usefulness of these appraisals. The goal of the present study was to examine the effectiveness of source monitoring training in reducing some of these biases, namely the halo effect and the negativity bias. One hundred twenty six participants watched the videotaped performance segments of an instructor that presumably took place over the course of a year. After watching the video, participants in the experimental group received source monitoring instructions teaching them how to use their memories. In particular, they were trained to differentiate specific memory traces from feelings of familiarity and base their judgments solely on these specific memory traces. Results showed that source monitoring training successfully reduced halo effect and negativity bias and increased the accuracy of ratings. These improvements were due to greater reliance on specific memory traces, as revealed by measures of recognition memory which indicates the accuracy of the participants in discriminating between the behaviors that occurred and did not occur. In conclusion, source monitoring training appears to be a promising method for increasing the accuracy of ratings in performance appraisal settings.

Keywords: Source monitoring, performance appraisal, bias, recognition, meta-cognition

ÖZET

Performans değerlendirme hem iş dünyasında hem de eğitim kurumlarında düzenli olarak uygulanan bir süreçtir. Ancak, performans değerlendirme sürecinde yaygın olarak görülen çeşitli değerlendirici hataları performans değerlendirmelerinin sonuçlarının kullanılabilirliğini düşürmektedir. Bu çalışmanın amacı kaynak monitör etme eğitiminin (source monitoring training) hale ve olumsuzluk etkisini düşürmede ne kadar etkili olduğunu incelemektir. Bu çalışma için hazırlanmış deneysel ortamda 126 katılımcı bir öğretim üyesinin yıl boyunca videoya çekilmiş çeşitli davranışlarını izlediler. Videoyu seyrettikten sonra deney grubundaki katılımcılar hafızalarını nasıl kullanacakları konusunda bilgi veren kaynak monitör etme eğitimi yönergelerini aldılar. Böylece aşinalık duygularıyla hafızalarındaki detaylı bilgileri ayırt etmeyi, ve yargılarını sadece hafızalarındaki spesifik olaylara dayandırmayı öğrendiler. Böylece aşinalık duyguları yerine hatırlayabildikleri spesifik davranış göstergelerine odaklandılar. Sonuçlar kaynak monitör etme eğitiminin hale ve negatiflik etkisini başarılı bir şekilde düşürdüğünü ve performans değerlendirmelerinin doğruluğunu arttırdığını gösterdi. Katılımcılara videoda gerçekleşen ve gerçekleşmeyen olaylar soruldu ve bu olayları videoda gözlemleyip gözlemlemedikleri tanıma ölçeği aracılığıyla soruldu. Tanıma ölçeğinin gösterdiğine göre eğitimi almış olan katılımcılar olayların olup olmadığını kontrol grubu katılımcılarına göre daha iyi ayırt ettiler. Performans değerlendirmelerindeki bu gelişim spesifik hafızalara dayandırılan yargılardan dolayı ortaya çıktı. Sonuçların gösterdiğine göre kaynak monitör etme eğitimi performans değerlendirme süreçleri için uygulamaya değer ve doğruluğu arttırıcı etkili bir yöntemdir.

Anahtar kelimeler: Kaynak monitör etme, performans değerlendirme, değerlendirici hatası, tanıma, üst biliş

DEDICATION

To Ayşe Belgün, Didem and Kadir Gökcel and Jerfi...

ACKNOWLEGMENTS

From the first day of my life in Koc University, I know that it will be hard but an amazing experience to be a graduate student after graduating from ITU engineering department. While finalizing this thesis, I know that I can not be successful without the helps of the people that I will mention below.

First of all, I would like to thank my thesis father, Asst. Prof. Tarcan Kumkale, who has given me the opportunity of presenting my creativity, ideas, concerns and suggestions. With all of his goodness, open mind and the scientist point of view, I felt always motivated. He has planted in me the idea of exploring this very important topic and encouraged me during all the stages. He has a real belief in me more than anyone else has even I am in a desperate mood. His wisdom made be able to continue working at all times. This thesis, I believe, could not be completed without his valuable comments and insights but also his patience throughout the semester, even at times when I have fallen behind our agreed schedule. For me, it was such an honour, unique opportunity and, indeed, privilege to be able to work with him after following his commentaries. He is the best advisor of all times for me.

I would also like to thank my committee members Prof. Canan Sümer and Asst. Prof. Mahmut Bayazit. I am grateful for their valuable comments and contributions to this thesis. They answered to all of my questions and provided their ideas; so that I feel we are scientists all trying to examine a new training together. I also want to thank Asst. Prof. Mahmut

Bayazit, Prof. Zeynep Aycan, Assoc. Prof. Nazlı Baydar and Prof. Sami Gülgöz for their teaching, mentoring and listening. I am also indebted to Assoc. Prof. Fatma Küskü and Asst. Prof. Mehmet Erçek for showing me the way and provide support.

This thesis would not have been realized without the support I have received from my friends from I/O and Developmental Psychology Departments. I have been blessed with many good friends who have been with me all along the way with their friendship and support and made my life on campus more enjoyable. I will always be grateful to especially to Selin Derya, Ayseul Ertüreten, Şirin Duygulu and Soner Dumani as the memories of our shared moments will stay with me for the rest of my life. Also, I am grateful for their eagerness to help during video construction and data collection process. I am also indebted to Begüm, Emre and Çağlar for their support in video construction. As a director of this short film constructed specifically for this study, I believe that I can not find any cameramen, actor and director of photography better than them.

I would also like to thank to TÜBİTAK and Koc University for financial support.

Finally, I would like to thank my family for their unconditional love and belief in me. My mother, Ayşe Belgün Gökcel always made me feel that I am valued, different, genius and beloved. I would not be able to succeed in the things without her love, belief and support. I would like to thank Jerfi Önal. During the hard times of the past 2 years, he held my hand and was always with me with all his love, courage and wisdom.

TABLE OF CONTENTS

STATEMENT OF AUTHORSHIP.....III

ABSTRACT.....IV

ÖZET.....V

LIST OF TABLES.....XII

INTRODUCTION.....1

LITERATURE REVIEW.....6

2.1. HALO EFFECT..... 6

2.2. NEGATIVITY BIAS IN PERFORMANCE APPRAISAL.....7

2.3. INTERVENTION STRATEGIES.....9

2.4. DUAL PROCESS OF MEMORY.....13

2.4.1. Remember and Know Judgments..... 14

2.4.2. Negative Impressions as a Source Monitoring Confusion..... 14

2.5. SOURCE MONITORING TRAINING.....16

2.5.1. Source Monitoring Training and the Halo Effect.....18

2.5.2. Source Monitoring Training and the Negativity Bias..... 21

2.6. THE PRESENT STUDY.....24

METHOD.....25

3.1. PARTICIPANTS AND DESIGN.....26

3.2. PROCEDURE..... 26

3.3. MATERIALS.....28

3.3.1. Stimulus Video.....28

3.3.2. Source Monitoring Training.....29

3.3.3. Performance Evaluation Form..... 30

3.3.4. Behavioral Questionnaire- Recognition Memory..... 31

3.4. DEPENDENT VARIABLES.....32

3.4.1. Halo.....32

3.4.2. Negativity Bias and the Rating Accuracy.....33

3.4.3. Recognition Accuracy.....33

3.4.4. Response Time.....34

RESULTS.....35

4.1. TESTING THE EFFECT OF SOURCE MONITORING TRAINING ON PERFORMANCE APPRAISALS.....35

4.1.1. Negativity Bias as Reflected on Specific Ratings.....36

4.1.2. Negativity Bias and Rating Accuracy.....37

4.1.3. Effect of Source Monitoring Training in Reducing Halo Effect.....38

4.1.3.1. *Ratings on an Irrelevant Dimension*.....38

4.1.3.2. *Correlations*.....39

4.1.3.3. *Reliabilities on the Performance Appraisal Form*.....42

4.1.3.4. *Absolute Squared Deviations*.....43

4.1.4. Recognition Memory..... 43

4.1.4.1. *The Effects of Source Monitoring Training on Hit Rates and False Alarm Rates*.....44

4.1.4.2. *The Effects of Source Monitoring Training on Behavioral Accuracy*.....46

4.2. RESPONSE TIME DATA.....47

4.3. SUMMARY OF RESULTS.....48

DISCUSSION.....49

5.2. ORDER OF INFORMATION.....49

5.2. THE EFFECT OF SOURCE MONITORING TRAINING IN REDUCING NEGATIVITY BIAS.....50
5.3. THE EFFECT OF SOURCE MONITORING TRAINING IN REDUCING HALO EFFECT.....51
5.4. RECOGNITION MEMORY IMPROVEMENT WITH SOURCE MONITORING TRAINING.....52
5.3. CONTRIBUTIONS OF THE PRESENT STUDY.....54
5.4. LIMITATIONS OF THE PRESENT STUDY AND SUGGESTIONS FOR FUTURE RESEARCH.....54
REFERENCES.....59
APPENDICES.....67
APPENDIX A.....67
APPENDIX B.....70
APPENDIX C.....72

LIST OF TABLES

4.1. Correlations among study variables in the control condition.....41

4.2. Correlations among study variables in the source monitoring condition.....41

4.3. Reliability coefficients.....42

4.4. Recognition memory measures across conditions.....45

Chapter 1

INTRODUCTION

Performance appraisal is one of the key human resources practices in work settings today. In a typical organization, managers evaluate the performance of their subordinates once or twice a year. Similarly, in educational settings, students and teachers evaluate each other on a regular basis. When appraising performance, raters are expected to make accurate and fair decisions. Decades of research investigating this topic, however, shows that rater biases are quite common (for a review, see e.g., Murphy & Cleveland, 1995; Dessler, 2005).

Several steps are involved in the performance appraisal process: The rater observes the behaviors of a particular ratee, forms a cognitive representation of these behaviors and stores these representations in memory. When the appraisal day comes, the rater retrieves stored information from memory and integrates this information with other information (Denisi, Cafferty, & Meglino, 1984). Given the cognitively taxing business environment, each one of these steps in the performance appraisal process can be disrupted at some point and bring about inaccurate appraisal decisions. An inaccurate appraisal can be detrimental because these decisions are used in many ways (Cleveland, Murphy, & Williams, 1989). For instance, managers use performance information in developing the workforce and making employment plans; employees use it in determining their career, training, and development plans. In sum, an inaccurate appraisal will not only demoralize the employee but will also affect other HR

activities. Therefore, researchers and practitioners have long been interested in developing strategies to increase the accuracy of raters and the system in general (Baltes & Parker, 2000).

Researchers studying performance appraisal typically focus on the measurement related aspects of the issue (for a review, see Murphy & Cleveland, 1995). Several errors and biases have been identified in this literature, including positive/negative leniency, confirmation and escalation bias, central tendency bias (e.g., Dessler, 2005), performance cue effects (e.g., Baltes & Parker, 2000), recency/primacy effects (e.g., Haugtvedt & Wegener, 1994), assimilation effects and contrast effects (e.g., Sümer & Knight, 1996), negativity biases (e.g., Skowronski & Carlston, 1989) and halo effects (e.g., Lance, LaPointe, & Stewart, 1994). Compared to dozens of studies demonstrating the ubiquity of these errors and biases, interventions to counteract them have received relatively scarce attention. One such promising intervention is the recently introduced source-monitoring training (Martell & Evans, 2005), which has been demonstrated to reduce the effect of initial impressions in behavioral measurement. Specifically, participants in the Martell and Evans (2005) study were induced to develop positive or negative expectations about the performance of a ratee prior to watching his actual performance. After watching this performance, some of the participants received a training in which they were taught to distinguish memories for specific incidents from feelings of familiarity. The authors expected that if participants could discount feelings of familiarity and instead rely on the specific instances that they could recall from the performance video, the effects of prior expectations would be reduced. They provided some instructions. For example, they asked participants to think of a memory like seeing someone on the street whom they met at an occasion. They described that if the participants had a mental image of seeing the person, the memory of speaking with the person or brought to their minds an associated event making them recognize those event, the participants had a

remember judgment. If the participants confidently recognize that they have seen the person on the street without bringing the memory in mind and reliving the event, they had feelings of familiarity that formed know judgments. The results confirmed source monitoring training reduced the biasing effects of initial performance expectations. Thus far, the effectiveness of this training has been explored only in this study. This new approach was not investigated in a performance appraisal setting. We think that exploring the idea of source monitoring training can be beneficial in performance appraisal settings since raters demand on their memory and benefit from behavioral information. Source monitoring training may be a useful approach in reducing rater biases specifically the ones result from insufficient rater memories. Therefore investigating the effect of this novel approach in eliminating rater biases is worthwhile. Moreover, questions remain regarding whether or not this intervention can be applied to reduce different rater biases in performance appraisal. The goal of the present study was to address this question by examining the effect of source monitoring training in reducing halo effect and negativity bias by way of increasing accuracy.

Halo error is an error that was investigated in the domain of social cognition by generations of researchers (Goffin, Jelley, & Wagner, 2003). Lance et al. (1994) reviewed different mechanisms for halo error. General impression model of halo error defines the tendency to rate a ratee according to a global evaluation (Lance et al., 1994; Nisbett & Wilson, 1977). For example, one manager's overall evaluation of an employee can reflect on other performance dimensions. Salient dimension model of halo error defines halo as an error occurs when a salient dimension affects the ratings on the other dimensions. For example, assuming that planning and organization can be a salient performance dimension, if an employee is superior in planning and organization, she/can be classified as a better employee compared to a situation that she/he is superior in communication. Inadequate discrimination

model of halo error (e.g., Murphy & Reynolds, 1988; also see Lance et al., 1994) defines the source of halo as the failure to distinguish among different dimensions of performance. For example, evaluation of a trait such as friendliness, may bias the ratings of specific work-related behaviors such as planning and organization. In other words, halo effect emerges because raters rely on an existing evaluation, overall impressions, or information about a seemingly relevant other dimension, instead of the information exactly relevant to the target dimension. Because the source monitoring training is targeted at identifying and using memory traces for specific incidents relevant to the questioned dimension, it is very likely to be useful in reducing halo effects in performance appraisal settings.

In general, halo effect results in inflated employee ratings. Sometimes, however, some of the same mechanisms governing the halo effect can also generate an opposition pattern colored by inflated negativity. For instance, a rater may evaluate an employee as being deficient in a critical dimension or trait, and then consciously or unconsciously assume that the person is deficient in other dimensions or traits as well. For example, an instructor who is judged to be poor in his or her interactions with students in class may be judged to be deficient in his or her enthusiasm, fairness in grading of assignments, or organizations skills as well. Sometimes, this phenomenon is described as the *devil effect* (Thorndike, 1920); but I will use the more general term *negativity bias* in referring to such instances (e.g., Skowronski & Carlston, 1989).

Negativity bias occurs when a specific rater is under the effect of negative behaviors of the ratee and overlooks the positive behaviors of the ratee. In many settings, negative behaviors may be seen as more diagnostic of the performance than positive behaviors; and hence negative behaviors may have a disproportionately greater influence on ratings than

positive information (Ito & Cacioppo, 2005; Smith, Larsen, Chartrand, Cacioppo, Katafiasz, & Moran, 2006). Thus, people may think that a specific person is likely to have other negative attributes if he or she has already one. Participants under the effect of negativity bias tend to ignore or underestimate instances of positive performances, while remembering and using negative instances to a great extent. Presumably, they do not search their memory for positive instances as deeply as they do for negative instances. Thus, any intervention encouraging raters to look for specific memory traces may reduce negativity bias. In that regard, source-monitoring training should be especially useful because it revolves around teaching individuals how to use their memories in conducting ratings.

In sum, source-monitoring training is a potentially useful technique for reducing halo effects and the associated negativity biases. Thus far, the effects of source-monitoring training in reducing halo effect and negativity bias have not been explored. This study aims to examine the effectiveness of source-monitoring training in reducing these biases in a performance appraisal setting. In the following sections, I will begin by first reviewing research on halo effect and negativity bias. Then, I will review previous attempts to reduce these effects. Finally, building on dual-process models of memory, I will argue that source-monitoring training intervention can be effective in reducing halo effect and negativity bias in performance appraisal settings.

Chapter 2

LITERATURE REVIEW

2.1. Halo Effect

Halo effect takes place typically when raters simply recall the readily accessible general impression they formed about the ratee and use that impression for making their judgments (Denisi & Peters, 1996). When raters are under the effect of their overall impressions halo effects emerge more readily.

In addition to the three models of halo defined by Lance et al. (1994), Balzer and Sulsky (1992) summarized aspects of halo error in performance appraisal; "general impression halo" and "dimensional similarity halo." General impression halo results from the retrieval and use of an overall evaluative impression in making an appraisal. Overall evaluations can be stored separately from other relevant information, and raters may use these overall impressions instead of recalling specific details that constituted the foundations of those overall impressions (Feldman, 1981). This effect can occur during encoding or retrieval stage. When raters pay specific attention to behaviors congruent with the general impression and underestimate crucial but incongruent behaviors, halo effect will occur during encoding. Thus, when a general impression is readily available, raters may choose to go by these impressions instead of choosing the more challenging route of retrieving specific information from memory and computing a judgment based on careful consideration of the retrieved information. In that case, specific ratings along various dimensions of performance will

probably depend on this overall impression (Palmer, Thomas, & Maurer, 2003). Furthermore, correlations among performance dimensions will be artificially inflated (Goffin et al., 2003).

It is likely that the general impression halo is quite common in both educational settings and work organizations. Critical performances take place throughout the year but the decisions are made once or twice a year. Due to the cognitively and physically taxing pace of life in schools and work organizations, students and managers will rarely possess sufficient ability and motivation to retrieve everything that they know about the ratees. Instead, they will most likely retrieve an overall impression and perhaps a few bits of specific information along with that impression and base their appraisals exclusively on these grounds. Recalling behaviors of the ratee according to each specific performance indicator should be quite difficult for most raters.

As for "dimensional similarity halo", it occurs when a behavior or attribute is used to form two or more trait performance dimension impressions or when performance dimension impression is used to make ratings regarding a dimension for which there is little or no relevant information (Balzer & Sulsky, 1992). For example, it may be difficult to differentiate skills and behavioral performances related to planning and from skills and performances related to organization. Therefore, ratings of an employee on these two dimensions can be artificially inflated (Feldman, 1986).

2.2. Negativity Bias in Performance Appraisal

Previous studies showed that negative information may have a disproportionately greater impact on performance decisions than positive information (e.g., Oden & Anderson, 1971; Wojciszke, Brycz, & Borkenau, 1993). Similar findings were reported in attitude

formation and person perception research (for a review, see Skowronski & Carlston, 1989). Baumeister, Bratslavsky, Finkenauer, and Vohs (2001), for instance, reported that bad impressions and stereotypes have a greater impact on final impressions than positive impressions, and are more resistant to disconfirmation than good impressions and stereotypes. In addition, the consequences of bad events will be more permanent and intense than the consequences of good events. Similarly, Pratto and John (1991) reported that memories for bad traits in person perception tend to remain more accessible over time than memories for good traits (also see Ikegami, 1993).

According to the averaging model of information integration (Anderson, 1981) impressions of raters towards a stimulus person should be neutral if the person depicts the same number of positive and negative behaviors of similar absolute valence. Thus, the final impression of a target person should be average of the true scores of behaviors. It is not what the empirical tests of the model finds, however; even when the number and extent of positive and negative behaviors are balanced, negativity bias prevails (e.g., Fiske, 1980; Oden & Anderson, 1971). Thus, as a consequence of negativity bias a person can be negatively rated and given ratings lower than his/her true level of performance.

According to Martell and Willis (1993), two mechanisms are especially relevant in the context of how initial impressions distort behavioral ratings. The first mechanism occurs when initial impressions serve as a type of perceptual filter directing attention to congruent information (Johnson & Judd, 1983; Srull & Wyer, 1983; White & Carlston, 1983). Thus, a rater observes the ratee on the basis of initial impression. For example, Anderson and Slusher (1987) reported that once a ratee is categorized as a poor performer, raters may overestimate the frequency of behavioral incidents consistent with that impression. As a second

mechanism, initial impressions may cause discrepant performance information to be assimilated into the existing impression of the ratee, suggesting that initial impressions may affect encoding and storage of the information. Murphy and Balzer (1986) stated that accuracy in performance appraisals depends in part on the rater's ability to accurately remember the behavior of the individual he or she is evaluating. People search information that matches their existing attitudes (consistent information) and try to avoid information that is incomparable with their attitudes (consistent). Encoding mechanisms work more effectively for negative stimuli commanding attention to encode their significant properties (Ochsner, 2000). Thus, negativity bias is frequently observed.

In short, negative impressions influence ratings much more so than positive information. For example, Thorndike (1920) defined a phenomenon called "devil effect" in which a rater observing one bad behavior about a ratee believes that the ratee must have other negative behaviors or characteristics. In that regard, the devil effect is closely related to the halo effect; indeed, it would not be wrong to suggest that it is similar to the halo effect, as the mechanisms generating the two effects are generally the same.

2.3. Intervention Strategies

There are various interventions designed in an attempt to reduce biases in performance appraisal. In a review of these interventions, Woehr and Huffcutt (1994) reported that considerable attention has been devoted to rater training for improving performance rating decisions. The authors classified rater-training strategies into four groups: Rater error training, performance dimension training, frame-of-reference training and behavioral observation training.

In a typical rater-error training, raters receive plain information about the most frequently observed rating errors such as halo and contrast errors; it is assumed that making raters familiar with these errors and biases would make them more sensitive and thus accurate Woehr, D. J., & Huffcutt, A. I. (1994). On the other hand, performance-dimension training has a different focus; the goal of this training is to give raters an understanding of the appropriate dimensions on which ratings should be based. Frame-of-reference training defines performance dimensions and provides examples of behavioral incidents according to each dimension (e.g., Noonan & Sulsky, 2001). In other words, frame-of-reference training makes raters understand what behaviors comprise specific levels of performance based on specific dimensions and present raters performance samples that prevent raters from memory loss (Roch & O'Sullivan, 2003). Finally, behavioral observation training focuses on strategies to improve the observation and recording process of behavioral information (Woehr & Huffcutt, 1994). The most significant output of the review (Woehr & Huffcutt, 1994) is that the effectiveness of these intervention strategies varies as a function of outcomes against which effectiveness is judged. For instance, rater-error training tends to decrease errors in ratings, whereas performance-dimension training and behavioral-observation training (i.e. diary keeping) typically increases observational accuracy; frame-of-reference training increases rating accuracy. In sum, further research on interventions designed to increase the accuracy of both ratings and memory are still needed.

Frame-of-reference training (FOR) was found to be effective in reducing halo error (Woehr & Huffcutt, 1994). However, research showed that FOR training caused raters report behaviors that were not exhibited by the ratees (Sulsky & Day, 1992). Noonan and Sulsky (2001) interpreted this finding that even though FOR training facilitates the construction of correct impressions about ratee performance, these impressions may make raters incorrectly

recognize behaviors which are consistent with their impressions. It would seem, therefore, that further investigations are needed to increase recognition accuracy; so that raters are able to give concrete and specific feedback about the ratees' performances (Noonan & Sulsky, 2001). Woehr and Feldman (1993) reported that intervention strategies that raise raters' reliance on their memory for performance related information and minimize their use of heuristics when making behavioral ratings is not sufficient.

Raters observe the ratees in a specific period and try to retrieve behavioral information during the rating task. Thus, there are various interventions focusing on the improvement of memory. A successful rater training trying to reduce the biasing effects should increase the accessibility and diagnosticity of ratee behavioral information. The accessibility and diagnosticity of information in memory may affect evaluations (Feldman & Lynch, 1988). Accessibility refers to the easiness with which a cognitive construct can be brought into awareness, whereas diagnosticity determines whether or not previous evaluations or stored information are perceived to be relevant to following evaluations. In other words, accessibility and diagnosticity influence whether a prior cognition will be used as an input to a related judgment (Bauer & Baltes, 2002). According to Baltes and Parker (2000), accessibility of these specific memories and the possibility of using these memories at the time of rating can be increased if raters recall behaviors that were depicted by the ratee before the performance appraisal. Specifically, they asked raters to recall as many positive and negative performance behaviors as possible before completing the evaluation and then encouraged raters to use these recalled behaviors to complete their ratings. Findings showed that a structured free recall intervention reduced a rater's reliance on an overall judgment of the ratee (which is often biased by initial expectations) by getting the rater to use specific observed behaviors to complete the performance ratings (Baltes & Parker, 2000).

Structured free recall intervention increases the accessibility of behavioral information and decreases raters' reliance on heuristics. Consequently, it eliminates the biasing effects of rater expectations (Bauer & Baltes, 2002). Different interventions can be designed to reduce the biasing effects from performance appraisal by increasing accessibility of relevant behavioral information.

In parallel, in a recent study introducing a new training approach, Martell and Evans (2005) gave participants instructions about remember and know judgments which are investigated under dual process of memory. Specifically, they asked participants to use remember judgments which are based on specific memory traces and asked participants not to use know judgments which can be affected by their feelings of familiarity by conducting ratings. Results showed that the participants who received source monitoring training were prevented from the biasing effects of performance cues (Martell & Evans, 2005). In the study, training participants were asked to rely on their specific memory traces after watching a stimulus video. Thus, they used behaviors which they actually picture and relive while conducting ratings. Since source monitoring training asks raters search their memory and report only specific behaviors, it may increase the accessibility of the relevant behavioral information. Consequently it is a promising method for increasing the accuracy of performance appraisal decisions.

Remember and know judgments and the difference between them formed the main idea of source monitoring training (Martell & Evans, 2005). By discriminating between these judgments, the accessibility of the behavioral information was increased. Besides, the participants used remember judgments and were not affected by the know judgments which

can bias their appraisals and feedbacks. Remember judgments refer to recollection of prior behaviors that are typically unaffected by firstly presented information, whereas know judgments refer to feelings of familiarity that are typically affected by the firstly presented information. These kinds of judgments were investigated under dual process of memory. Thus, in order to dwell upon the application of source monitoring training in performance appraisal, I will first review research on dual-process of memory.

2.4. Dual Process of Memory

Most memory researchers concur that the personal experience of recognition is composed of two mechanisms: Remembering and knowing (e.g., Gardiner & Java, 1993). Gardiner (1988), for instance, defined remembering as the ability to be aware of different characteristics of an event or other things associated with the event and knowing as the recognition that the event occurred without consciously recollecting any information associated with the event. In addition, remember judgments stand on specific and detailed memory traces and they can be mentally experienced again whereas know judgments are the "feelings of familiarity" and take place when there is not any mental recollection of the memory itself (Martell & Evans, 2005). The authors provided a specific example in order to clarify the meaning of these two types of judgments. The situation of meeting a woman and trying to decide if she is recognized as an acquaintance is given. A person can have a remember judgment if he or she recognizes having met with the woman last week at a dinner party, recalling that her clothing was overgenerous and the chat with her was boring. On the other hand, if the person has no conscious recollection regarding that woman and a familiarity feeling makes the person recognize her as an acquaintance with the same level of confidence, the person has a know judgment.

2.4.1. Remember and Know Judgments

Whether people can distinguish remember judgments from know judgments is worth considering: Tulving (1985) reported preliminary evidence suggesting that remembering and knowing can be experienced as two different states of awareness during retrieval and people can distinguish between these states. In line with this, Martell and Evans (2005) trained raters to distinguish between remember and know judgments and to report only behaviors that evoked remember judgments which are detailed memories and repress reporting behaviors that are based on know judgments which are kind of feelings of familiarity.

2.4.2. Negative Impressions as a Source Monitoring Confusion

In contemporary theorizing, the *source* of a representation refers to a variety of characteristics that identify the conditions such as the spatial, temporal and social context of the event under which a memory is acquired (Johnson et al., 1993). The source-monitoring framework mainly refers to differentiating memories for thoughts and imaginations from memories for perceived items (Johnson & Raye, 1981). In other words, source monitoring refers to the set of processes involved in making attributions about the origins of memories, knowledge and beliefs (e.g., Hashtroudi, Johnson, & Chrosniak, 1989; Kahan & Johnson, 1992).

Source-monitoring tasks involve asking people to make attributions about the origins of their memories. For instance, people may be asked to make a decision about who said what or to identify whether the information was seen in a picture or read in a document (Hekkanen

& McEvoy, 2002). The authors suggested that if people attribute the memories to a wrong source, they make source misattribution error. Thus, source misattributions emerge from chronic or temporary failures to determine the origins of cognitions. Source misattributions may result in inaccurate performance appraisal and feedback.

If a specific ratee's negative performance information is highly accessible in memory, this information may bias rater's performance appraisal decisions (Baltes & Parker, 2000) by leading the rater to adopt a biased decision criterion at the time of making a judgment, such that the rater becomes more willing to recognize behaviors that are congruent with his or her expectations during judgment construction. It is also possible that such performance expectancies may affect judgments by way of distorting perception or encoding, but the available evidence did not provide any support for this possibility (e.g., Martell &Willis, 1993; Murphy & Jones, 1993).

Similar to negative impressions, stereotypes provide expectations about what kinds of behavior may be expected from individuals (Hamilton & Sherman, 1994; Sherman & Bessenoff, 1999). Thus, Sherman and Bessenoff (1999) reasoned that stereotypes act as heuristic cues that influence the decision criteria for making a source attribution. In their study, they gave participants a list of behaviors including friendly, unfriendly and neutral behaviors. Participants were asked to memorize the behaviors on this list. In a second list, a specific person is portrayed with the help of friendly and unfriendly behaviors and he was described as either a skinhead or a priest. For participants who were said that the person was a skinhead, unfriendly behaviors were stereotype-consistent and friendly behaviors were stereotype-inconsistent. For participants in the priest condition, the opposite was the case. In the second session a third list including thirty friendly and unfriendly behaviors from the

previous lists and thirty new behaviors were presented to the participants. Results showed that participants who were given additional distracting tasks while making source attributions in the recognition task (low capacity condition) used stereotypes to provide cues whether the behaviors occurred or did not. Thus, they made many more source misattributions for stereotype-consistent than stereotype-inconsistent behaviors.

In sum, previously formed judgments and stereotypes may cause source monitoring confusion. Raters can search for information that is consistent with their judgments and ignore inconsistent information with their initial judgments. Consequently, they may not distinguish between observed and unobserved ratee behaviors accurately at the time of the rating.

2.5. Source Monitoring Training

Source monitoring training is an extension of reality monitoring (Martell & Evans, 2005). Source monitoring is an effective approach in which individuals are trained to distinguish between behaviors exhibited by the ratees' and behaviors that are not exhibited by the ratees' but are consistent with raters' implicit theories (Johnson & Raye, 1981). In line with this, Martell and Willis (1993) suggested that asking raters simply to memorize the ratee performance information (reality monitoring strategy) may be effective in reducing the biasing effects of rater's expectations. However, they could not find a relationship between memories of performance related information and rater's expectations showing that the rating bias was not occurred while raters observe ratee behaviors. The authors suggested that since performance expectations effects arise from "inferential bias" at the time of the rating, trainings should be applied just before the rating task. Source-monitoring training fulfills this need as it is introduced to the raters at the time of rating.

Source monitoring can rely on both systematic and heuristic processes at the same time (Sherman & Bessenoff, 1999). Systematic source monitoring is effortful since it requires detailed examination of the relevant information in order to find out the possibility that the information is coming from various sources. On the other hand, heuristic source monitoring depends on effortless usage of simple cues when an individual attributes memories to the source. These simple cues can be feelings of familiarity or previous schemas and source expectancies (Johnson, Hastroudi, & Lindsay, 1993).

In sum, expectations can affect source monitoring decisions in a biased manner. When raters' capacity to process information is low, the possibility that raters rely on heuristic source monitoring processes and attribute stereotypical behaviors to the source is increased (Sherman & Bessenoff, 1999). In line with this, Cook, Marsh, & Hicks (2003) classified participants according to their capacity; some of them had full attention some of them had divided attention. Participants who had divided attention made biased attributions almost twice as larger than the participants who had full attention. Therefore, because stereotypes such as initial negative impressions, expectations and overall evaluations may cause raters to make misattributions especially when the capacity to process information is low, increased ability to make systematic source monitoring is needed. Source-monitoring training requires that the raters engage in systematic source monitoring; consequently, it should be a useful approach in reducing rater biases mentioned before.

Source monitoring in performance appraisal refers to a specific rater's attributions about his or her memory for the ratee behaviors. An accurate source monitoring leads to accurate performance appraisals by making raters distinguish the actual ratee behaviors from the behaviors that did not occur in reality. If raters cannot distinguish behaviors that occurred

and that did not occur, source misattributions will most likely lead to inaccurate performance appraisals. Therefore, source-monitoring training gives raters information about how to distinguish remember judgments from know judgments and use remember judgments while making ratings.

Source monitoring training has not been used in a performance appraisal setting before. Nonetheless, there are some studies that can shed light on to its potential effects. For instance, source monitoring training has been shown to increase recognition accuracy. Thierry and Spence (2002) conducted a study in order to see whether a source monitoring training procedure would decrease 3-4-year old children's suggestibility by enhancing their ability to detect and reject false information. Children who were prompted to report source information and detect misleading questions regarding source information by source monitoring training were more accurate in recognizing and correcting source misattributions than children who were not given source-monitoring training. In another study, Martell and Evans (2005) examined the effects of this training in reducing biases that arise from reliance on performance cues. As mentioned before, they found that source monitoring training reduced the effects of performance cues. In short, there is some evidence suggesting that the source-monitoring training should work in a performance appraisal setting as well.

2.5.1. Source Monitoring Training and the Halo Effect

Martell and Evans (2005) suggest that source monitoring training may be useful in increasing raters' ability to accurately categorize ratees according to performance levels—namely, classification accuracy (Sulsky & Day, 1992). Lord (1985) conceptualized classification accuracy in connection with an overall impression or classification of ratee's

performance. If classification accuracy can be increased by source monitoring training, halo effect may also be reduced since the accessibility of relevant information was increased.

People who have a tendency to make incorrect source attributions without spending effort to retrieve a specific behavior or judgment from their memory, tend to rely on know judgments that can be defined as a sense of familiarity and overall evaluation (Martell & Evans, 2005). Therefore, if raters are trained to distinguish sources of their judgments, they may be able to make more appropriate performance appraisals by way of decreased reliance on overall impressions.

In addition, if a memory response is experienced sufficiently and uniquely specifying a moment in one's past, it is experienced as remembering judgment (Bodner & Lindsay, 2003). Therefore authors stated that the possibility that people will remember something should depend in part on how fine they remember the answers to the other questions in a given situation. If raters search for their memory in order to distinguish their remember judgments from know judgments with the help of source monitoring training; because the information becomes more accessible in their memory, it would be easier to remember other behavioral incidences that will not be remembered otherwise. Judgments are related to overall impressions if memories of behavioral information are not relevant and easily accessible (e.g. Sulsky & Day, 1994; Sanchez & De La Torre, 1996). Increasing the accessibility of the information, source monitoring training is a promising method in eliminating halo effect. This way, raters will not solely rely on their overall ratings and will able to use other sorts of behavioral information.

In sum, raters use their overall evaluations to conduct performance appraisals unless they have relevant and easily accessible behavioral memories (Sulsky & Day, 1994; Woehr, 1994). Woehr and Feldman (1993) found that increasing the accessibility of specific performance relevant behaviors result in greater dependence on the behavioral information which leads to increased accuracy. In addition, research found that judgments are related to behavioral memory traces when such memories are relevant and easily accessible and to overall evaluations otherwise (e.g., Sanchez & De La Torre, 1996). Raters are able to attribute behaviors in each performance dimension accurately with the help of source monitoring training. Besides, by the help of the instructions of this training which increases the accessibility of behavioral information by asking raters search their memory for behaviors and use solely remember judgments based on specific behaviors, halo may be reduced by source monitoring training. Therefore I will test the following hypothesis:

Hypothesis 1: Source monitoring training will decrease halo effect.

Halo effect and negativity bias are naturally closely related because they are likely to be guided by some of the same mechanisms, such as suboptimal memory search and reaching a judgment prematurely. For instance, raters can form negative impressions towards a ratee and evaluate the subsequent performances of the ratee in line with this impression, even if there may be some positive performances in this period. It is possible to see this case as an instance of negativity bias or an instance of halo effect. As can be seen, halo effect and the negativity bias are closely related. Nonetheless, for presentational purposes, I will address the halo effect and the negativity bias separately.

2.5.2. Source Monitoring Training and the Negativity Bias

According to Feldman (1981), raters may adopt behavioral information consistent with their expectations more readily than inconsistent information. When a rater fails to modify judgments in light of new behavioral evidence, performance appraisals can hardly be accurate and beneficial, which should be especially likely when the behavior in question is negative. Manzoni and Barsoux (1998) defined a new term "set-up-to fail syndrome" in order to present the effects of negative impressions. When this syndrome occurs in organizations, the manager begins to worry that the employee performance is very weak. Sensing these concerns, in turn, the employee starts questioning his or her own ideas and skills and loses motivation to take initiative. In that regard, the set-up-to fail syndrome is akin to self-fulfilling prophecies. The authors believe that this situation can be eliminated if the managers distinguish their emotions from reality and monitor their own reasoning and try to find out how often and to what extent the employee showed poor performance. Therefore, it is clear that source monitoring training, by making the raters separate remember judgments (that evoke detailed memories) from know judgments (feelings of familiarity) and use solely remember judgments in their decisions, will enable the raters to make accurate and fair decisions about the employees.

Negativity bias results from the fact that raters pay attention to specific negative behaviors depicted during the appraisal period and conduct performance appraisal on the basis of this knowledge. For instance, upon making raters believe a target person performed well vs. poor at the beginning of the study, Martell et al. (1995) found that being primed by the information, raters who believe a target performed well (vs. poorly) attributed more effective (and fewer ineffective) behaviors to that target regardless of the behaviors' occurrences.

Similarly, in a study participants received a list of behaviors and asked to answer the question "How well does each item fit your image of a leader?" Results showed that participants rated behaviors congruent with implicit theories of effective leaders (high prototypic leader behaviors) more quickly than low prototypic leader behaviors (Lord, Foti & DeVader, 1984).

The aforementioned evidence suggests that the effect of negative information may be reduced if the accessibility and diagnosticity of the positive information can be increased. It is likely that negative information is perceived to be more diagnostic than positive information because of its expectancy-violating nature (Skowronski & Carlston, 1989). For instance, students would generally expect their instructor to be organized while possessing the state of the art knowledge. Thus, students would not be particularly impressed if their instructors possessed these attributes. In other words, the presence of these attributes will not be that diagnostic about the performance of the instructor. However, the absence of these attributes will be diagnostic because it will reveal a lot about the performance of the instructor. In line with this, Smith et al. (2006) reported that negative stimuli elicit more attention compared to positive information, which indicates attention bias. Their study showed that attention bias to negative information was eliminated when positive stimuli were made accessible in memory by presenting it more frequently. The authors suggested that an intervention designed to increase the relative accessibility of positive information in memory may reduce negativity bias. In that regard, source monitoring training may be useful because it induces raters to search their memory deeply for all sorts of behavioral information.

A particular rater's inability to remember ratee behavior can be an obstacle to accuracy. Source monitoring training may remove this obstacle. In a study examining order

effects, Haugtvedt and Wegener (1994) manipulated the sequence of information about nuclear power plants. Half of the participants received information in pro/con order thus they received arguments that support nuclear power plants followed by arguments that were against nuclear power plants. Other half of the participants received the information in the opposite order. All of the participants were asked to indicate their attitudes about the nuclear power plants and recall as many arguments as possible. Results showed that participants who received the information in pro/con order gave more favorable ratings than the other participants who received information in the opposite order indicating effects of negative priming. This indicates that recall of the most accessible negative message may result in negativity bias. Source monitoring training asks raters to think of past behavioral information, conduct deep memory search, find remember judgments by way of avoiding feelings of familiarity. Thus, it may be a promising method for not only decreasing negativity bias also increasing rating accuracy.

In conclusion, one can reduce this effect by increasing the accessibility of not only initial but also most recent information with source monitoring training. Therefore; participants may be more accurate in terms of their memory of behavioral information.

Hypothesis 2: Source monitoring training will increase behavioral accuracy.

Source monitoring training can help reducing rater biases therefore increasing rating accuracy. Thus; negativity bias may be reduced. Therefore I will test the following hypothesis:

Hypothesis 3: Source monitoring training will decrease negativity bias

2.6. The Present Study

The present study addresses the issue of whether source monitoring training can be used to reduce halo effects and negativity bias in performance appraisal. In this study, participants watched a video depicting an instructor performing some behaviors presumably taking place over the course of a year. Prior to observing the video, all of the participants were informed about the performance appraisal goal. Thus, they were asked to role play a student of the instructor and evaluate the instructor in the video according to specific performance indicating dimensions. In settings where people are exposed to both positive and negative traits and behaviors, it is necessary to manipulate the order of information presentation. Thus, half of the participants first received the positive information and then the negative information, whereas this order was reversed for the other half.

After watching the video, experimental group participants received source monitoring training instructions whereas control group participants completed filler tasks. Participants in the source monitoring training condition were trained to distinguish between remember and know judgments and asked to use solely remember judgments at the time of completing the evaluation form and behavioral questionnaire. The source monitoring training procedure was similar to the procedure used by Martell and Evans (2005). Control group went over the same process without any interventions; thus they were not trained to distinguish between remember and know judgments and asked to rely on remember judgments while completing the evaluation forms.

Afterwards, participants in all conditions completed performance evaluation forms. After filling out performance evaluation forms, participants completed a behavioral instrument similar to the ones used in previous studies (e.g., Martell & Evans, 2005; Sanchez & De La Torre, 1996). The behavioral questionnaire, including forty behaviors that occurred and did not occur, was used to examine how negative impressions affect raters' recognition memory. Of the twenty behaviors that were included in the videotape, ten of them were positive and ten were negative. Of the remaining twenty behaviors that did not occur, half of them were positive and the other half were negative. Based on these responses, signal detection measures of accuracy such as hit and false alarm rates were computed.

Using these measures to measure recognition accuracy, I will test the following hypotheses:

Hypothesis 4: Source monitoring training will increase hit rates for both positive and negative information. (Given the prevalence of negativity in the control condition, hit rates for negative information will be artificially inflated in the control condition; thus, the effectiveness of the intervention may not be reflected on the hit rates for negative behaviors).

Hypothesis 5: Source monitoring training will decrease false alarm rates for both positive and negative information (Given the prevalence of negativity in the control condition, however, false alarm rates for positive information will be artificially low in the control condition; thus, the effectiveness of the intervention may not be reflected on false alarms for positive behaviors).

Chapter 3

METHOD

3.1. Participants and Design

Sixty male and 66 female Introduction to Psychology students from various departments of Koç University in Istanbul participated in the study in exchange for extra course credit. Sixty-eight percent of them were freshman, 16 percent of them were sophomore, nine percent of them were junior, and seven percent of them were senior. Participants were randomly assigned to the conditions of a 2 x 2 factorial design, in which Training (source monitoring training vs. no-training control) and Information Sequence (positive-negative, negative-positive) were the between-subjects factors.

3.2. Procedure

Participants were seated in computer labs in groups of 10 to 15. Participants watched a video depicting the instructor's different behaviors, projected on the whiteboard. The order of the performance information was manipulated such that half of the participants first watched the predominantly poor performance and then the predominantly good performance; whereas the other half watched the good performance first. Thus, half of the behaviors displayed had negative implications for the impending performance appraisal (poor performance), whereas the other half had favorable implications (good performance). To make the performance

appraisal task as similar as the task faced by appraisers in everyday life, we induced mild cognitive load on the participants as they watched the video. Specifically, we stopped the video in every seven minutes and instructed participants to work on an unrelated task for two minutes. For instance, in the first break, participants provided their attitudes towards some logos. Additionally, they performed some arithmetical calculations with the expectation of retaining the results in memory for subsequent tasks. Similar tasks have been used successfully in past research to increase the cognitive load (e.g., Ambady & Grey, 2002). Overall, the video was stopped and participants repeated these tasks three times during the observation period.

After watching the videotape, participants in the source-monitoring training condition received instructions about how to distinguish remember judgments from know judgments. Source-monitoring instructions were the same instructions used by Martell and Evans (2005) (see Appendix A). These instructions provided participants examples of remember and know judgments, trained them to discriminate between these judgments and made them use remember judgments. In addition to these instructions, participants received several examples from everyday life where this distinction becomes relevant. All of these materials were presented on the whiteboard using slides, as it is the usual practice in everyday settings.

While participants in the source-monitoring condition received this training, participants in the control condition completed filler tasks before the rating task. Specifically, they completed various individual difference measures assessing constructs such as need for closure (Webster & Kruglanski, 1994), need for cognition (Cacioppo**,** Petty, Feinstein, & Jarvis, 1996) and judgmental self-doubt (Mirels, Greblo, & Dean, 2002). Afterwards, participants in all conditions completed the performance evaluation form presented on the

computer using a program called the Media Lab. Using performance appraisal form, participants first rated the overall performance of the instructor, and then evaluated him along various other specific performance dimensions. After completing the performance appraisal task, participants completed a behavioral questionnaire designed to measure recognition memory (e.g., false alarms, hit rates). On this questionnaire, participants indicated whether or not the instructor performed a given behavior. As a final task, participants rated the instructor for one more time. Different than the control group, source monitoring training participants completed the individual difference scales after all of the tasks. Throughout the experiment, participants were not allowed to interact with each other.

3.3. Materials

3.3.1. Stimulus Video

The instructor in the video was played by a professional male actor hired for the present study. The actor performed the role of the instructor across various typical classroom situations. To make it as realistic as possible, he changed accessories and clothes across the scenes. The whole video featured 41 performance incidents describing classroom behaviors taken from a set of 250 previously scaled incidents of college teaching behavior (Sauser et al., 1979). In order to determine an effectiveness score, Sauser, Evans, & Champion (1979) asked students to record critical incidents of instructor performance that corresponded to five dimensions and determined an effectiveness score based on independent ratings of 100 students. These dimensions of instructor performance were (a) ability to present material, (b) interest in course and material, (c) relationship with students, (d) reasonableness of workload, and (e) fairness of testing and grading.

Similar to the procedure applied by Borman (1978), after the completion of video construction including 41 different scenes, a group of 7 graduate students who served as subject matter experts were briefed on the nature and content of the scenes and given several opportunities to view the scenes. Subject matter experts rated each scene by using the same scale that the participants used. They used 7-point likert scales ranging from 1(*extremely bad*) to 7(*extremely good*). The ratings of subject matter experts were used to obtain true scores of each scene.

Using the ratings of the experts, average ratings and standard deviations of the ratings for each scene were calculated. The average of experts' ratings were used as true score measures (for similar procedures, see e.g., Jelley & Goffin, 2001; McIntyre, Smith, & Hassett, 1984; Sulsky & Day, 1992). Three scenes on which the subject matter experts displayed disagreement (*SD*s > 1) were not included in the final set of behaviors chosen for the video. Two additional scenes were excluded to arrange the true score of each performance indicating dimension to be "4" on a 7-point rating scale. The final version of the video included 36 scenes with six to eight behaviors related to each of the five dimensions, and was about 24 minutes in length. The average inter-rater standard deviation for the retained scenes was 0.73, recorded along 7-point rating scales. The interrater agreement on these scenes was sufficiently high, (i.e., intraclass $r = 0.97$).

3.3.2. Source Monitoring Training

In training participants to distinguish between remember and know judgments, we followed the previously validated instructions and procedures developed by Martell and

Evans (2005; see Appendix A). Thus, participants were told that remember and know judgments go along with strong feelings of recognition but these two judgments do not differ according to confidence or certainty. In addition, participants received instructions about how to make use of this distinction while completing the performance evaluation form and the behavioral questionnaire. Specifically, while completing the recognition memory task (i.e., behavioral questionnaire), participants were asked to indicate that the behaviors occurred for those behaviors they recognized on the basis of a remember judgment or indicate that the behaviors did not occur for those behaviors that were recognized on the basis of a know judgment. Participants in the source monitoring condition were able to see a brief description of remember and know judgments while they were completing the behavioral questionnaire. In the control condition, participants were asked to indicate that the behaviors occurred for the behaviors that were present in the video. Participants were asked to indicate that the behaviors did not occur for the behaviors that were not present in the video. They did not receive any other instructions.

3.3.3. Performance Evaluation Form

Participants in all conditions rated the overall performance of the instructor as well as his performance along six specific dimensions using 7-point likert type scales. Five of the specific dimensions pertained to behaviors depicted in the video (relevant dimensions), whereas the last one was not (irrelevant). The irrelevant dimension was included to be used in the assessment of halo error across conditions (for similar procedures, see Sanchez & De La Torre, 1996). This unrelated dimension addressed instructor's "ability to evaluate student progress and give feedback." This additional dimension provided ideas about the rating biases when raters are faced with unexpected situations. If the participants were under the effect of

their overall judgments, they would make inferences and give similar ratings to their overall ratings even there were not any behaviors related to that dimension in the videotape. Overall, with the overall evaluation questions asked two times, the performance evaluation form contained 14 items (see Appendix B).

3.3.4. Behavioral Questionnaire- Recognition Memory

Individuals may overestimate or underestimate the likelihood of a behavior taking place (e.g., Martell &.Willis, 1993). Therefore, it is customary to examine signal detection measures of recognition memory, such as hit rates, false alarm rates and memory strength (e.g., Snodgrass & Corwin, 1988). In order to calculate these measures, a behavioral questionnaire was administered to all of the participants (see Appendix C). There were 40 items in the behavioral questionnaire. The items were randomly ordered for each participant. Twenty of the items depicted behaviors that had occurred in the videotape. Of these 10 were positive and 10 were negative. Of the remaining 20 behaviors that did not occur in the videotape, 10 were positive and 10 were negative. Participants were given these 40 behaviors and asked about their opinion whether those behaviors occurred or not. Participants used 6-point likert scales ranging from 1*(extremely sure that the behavior did not occur)* to 6 (*extremely sure that the behavior occurred).* Responses as 1, 2 and 3 were recognized as a "No, the behavior did not occur" response whereas responses as 4, 5 and 6 were recognized as a "Yes, the behavior occurred" response (e.g., Martell, Guzzo, & Willis, 1995).

3.4. Dependent Variables

3.4.1. Halo

It is possible to examine the presence of the halo effect in a few ways (for a review, see Balzer & Sulsky, 1992). Therefore, I used multiple methods to address the halo effect. First, I compared the ratings of the experimental group with the control group on the irrelevant performance dimension. In presence of halo effect, raters can be expected to rate the instructor along this irrelevant dimension based on their overall impressions or their impressions on the other five relevant dimensions. Thus, ratings on this irrelevant dimension should display consistency with the rest of the ratings, and suggest a process of assimilation, which can be detected by analyzing intercorrelations. The average of the intercorrelations provides a stable and reasonable index of halo (e.g., Murphy & Jako, 1989). Inflated intercorrelations should in turn inflate the reliability coefficients on the performance appraisal form; thus, as another indicator, I examined these coefficients. In principle, halo effect implies a high reliability coefficient. If the source monitoring training is useful in reducing halo effect, then the reliability coefficient for the experimental condition should be lower than the coefficient for the control condition. Finally, I examined the average of the absolute squared deviations of each performance dimension rating from the mean rating across dimensions (e.g., Balzer & Sulsky, 1992; Palmer et al., 2003).

3.4.2. Negativity Bias and Rating Accuracy

Accuracy was assessed with the overall distance measure reported by McIntyre et al. (1984). This measure provides an index of how close the ratings of the participants are to the true scores generated by subject matter experts. Thus, deviations from the true score estimates suggest inaccuracy. Distance accuracy is the average absolute value of the deviation of the obtained ratings from the true scores (McIntyre et al., 1984). Thus, ratings lower than the true score can be considered as negatively biased. In other words, accuracy was also used as a negativity bias indicator.

3.4.3. Recognition Accuracy

In the behavioral questionnaire, half of the items described behaviors that occurred in the video whereas the other half described behaviors that did not occur. On this questionnaire, participants were instructed to indicate whether or not each one of the behaviors occurred in the video. These responses were used in the computation of standard measures of recognition accuracy such as hit rates and false alarms. A hit is defined as a yes response to a behavior that occurred in the video and hit rate is the probability of saying yes to a behavior that occurred. A false alarm is defined as a yes response to a behavior that did not occur in the video and a false alarm rate is the probability of saying yes to a behavior that did not occur. With the help of these hit and false alarm rates, memory strength and behavioral accuracy were calculated. Memory strength indicates whether the effect of impressions' is due to memory for consistent information (Martell & Willis, 1993). Memory strength (Pr) measure for negative behaviors was calculated by subtracting false alarm rates for negative behaviors

from hit rates for negative behaviors. Memory strength measure for positive behaviors was calculated by subtracting false alarm rates for positive behaviors from hit rates for positive behaviors. Hits for both positive and negative behaviors were summed and an overall hit rate was calculated. False alarms for positive and negative behavior were summed and an overall false alarm rate was calculated. Behavioral accuracy was calculated by subtracting false alarm rates for positive and negative behavior from hit rates for positive and negative behavior to measure the participants' memory for behavioral information.

3.4.4. Response Time

With the help of the experimentation program, the response times of all participants were calculated. Thus; we were able to measure the time taken by the participants to fill performance evaluation form and behavioral questionnaire in.

Chapter 4

RESULTS

The goal of this study was to test the hypothesis that receiving a source-monitoring training would decrease the magnitude of halo effects and negativity biases in a performance appraisal setting. To verify the likelihood of the negativity bias, it was necessary to manipulate the order of the behaviors available to the raters. Therefore, the order of presentation was manipulated and included as a factor in the analyses to be reported presently. ANOVA tests examining the effects of order along with the training factor (SMT training vs. control) did not involve a significant main effect of order or an interaction with the training factor on any of the measures of judgment and memory. In other words, the effect of the information order was not observed in the current study, presumably because of the forced exposure paradigm used in this study: Participants had to watch the video from the beginning to the end before reporting their evaluations. For simplicity of presentation, the effects of order will not be considered in the analyses reported in the following sections.

4.1. Testing the Effect of Source Monitoring Training on Performance Appraisals

Overall, there was evidence for negativity in this study. Where the true score of the overall instructor performance was 4 on a 7-point rating scale, the sample average implied strong negativity toward the instructor. All of the participants rated the overall instructor

performance twice; just after the video observation and at the end of the experiment. The first average overall rating (M = 2.60, SD = 1.24) and the second overall rating (M = 2.65, SD = 1.40) indicated negativity bias. The averages were under the true score point.

The first overall rating was requested before the completion of specific performance ratings; in other words, before participants found a chance to engage in cued-recall of specific incidents relevant to each dimension. Therefore, we did not expect a particularly strong effect of the training on the overall ratings reported before the ratings. Indeed, that was the case, F (1, 124) = 2.41, $p > .05$, d = 0.28. After the completion of the specific ratings, however, the expected effect of the training became evident. An ANOVA test confirmed that the ratings of the control group participants (M = 2.26, SD = 1.29) were more negative than the ratings of the training group participants (M = 3.02, SD = 1.40), F(1, 124) = 9.85, $p < .05$, d = 0.56.

4.1.1. Negativity Bias as Reflected on Specific Ratings

As for the effects of the training on ratings along specific performance dimensions, using the average of the ratings based on the five specific performance dimensions as the dependent variable, an ANOVA test confirmed that the ratings of the participants in the control condition were significantly more negative than the ratings of the participants in the training group, F(1, 124) = 9.70, $p < .01$, d= 0.56 . Thus, participants in the control condition found the performance of the instructor to be more negative (M = 2.29, SD = 1.12) than the participants in the training group (M = 2.87, SD = 0.97). Therefore, source monitoring training was useful in reducing the negativity bias. As described in the following section, however, the negativity bias was still pronounced.

4.1.2. Negativity Bias and Rating Accuracy

The findings presented above suggest that the ratings of the training group participants were closer to the true score estimates than the ratings of the control group participants. The true score estimates for all kinds of ratings was set to be 4 on a scale ranging from 1 to 7, where higher numbers indicated superior performance. Rating accuracy is typically assessed by computing an index of distance accuracy, which is the average absolute value of the deviation of the obtained ratings from the true scores (McIntyre et al., 1984). Using the overall distance measure for five performance dimensions as the dependent variable, an ANOVA test revealed that participants in the source monitoring training condition (M = 1.42, SD = 0.58) produced ratings that were significantly more accurate than those of participants in control condition (M = 1.92, SD = 0.68), F (1, 124) = 20.05, p<.001, d = 0.79.

In addition to the distance accuracy measure calculated for specific performance dimensions, distance accuracy of initial overall ratings and delayed overall ratings were also calculated. Using the distance accuracy of initial overall rating as the dependent variable, ANOVA was conducted. The main effect of the training was again significant, F(1,124) = 3.71, p <.05, d = 0.35. Thus, the initial overall evaluations of the participants who received the source monitoring training (M = 1.38, SD = 0.87) were closer to the true score point than the initial overall evaluations of those who did not receive the training (M = 1.75, SD = 1.16). When participants expressed their overall evaluations one more time after rating the instructor along various specific performance dimensions and filled in the behavioral questionnaire, the effect of the training in increasing distance accuracy was still pronounced, F(1,124) = 6.81, p <. 01, d = 0.46. Specifically, the ratings of those who received the training (M = 1.38, SD = 0.99) was still closer to the true score estimates than the ratings of those who did not receive

the training (M = 1.86, SD = 1.08). Overall, these data suggest that the source monitoring training successfully reduced the negativity bias by way of increasing accuracy of ratings. Overall, these findings support the hypothesis that receiving source monitoring would decrease the magnitude of negativity bias (see Hypothesis 3).

Next, I examined the question of whether the negativity bias observed in the study was guided by some of the same mechanisms that guide the halo effect. As mentioned before, the negativity bias can be considered as a corollary to the halo effect. Thus, an intervention decreasing the negativity bias could be expected to reduce the extent of halo effect as well.

4.1.3. Effect of Source Monitoring Training in Reducing Halo Effect

The presence of halo effect can be inferred from inflated correlations between the irrelevant dimension ratings and the relevant performance ratings. Thus, analyzing the reliability of a scale with relevant and irrelevant dimension ratings should not signal out the irrelevant dimension in the control condition where a strong halo effect is expected, but signal it out in the training condition. An alternative way to assess the halo effect may be to analyze the correlations between overall ratings and the irrelevant dimension ratings. In the control condition, this correlation should be stronger than the one observed in the training condition.

4.1.3.1. Ratings on an Irrelevant Dimension

In the video, the instructor displayed behaviors on five specific performance dimensions. The instructor did not depict any behaviors related to the dimension “ability to evaluate student progress and give feedback.” However, both the performance appraisal form

and the behavioral questionnaire included questions designed to address ratings and memory on this inapplicable, irrelevant dimension. Responses on this dimension could be used as a halo indicator. Specifically, halo effect is said to be observed if the ratings towards the actual specific performance dimensions were reflected on the ratings based on irrelevant dimension.

The results of an ANOVA test confirmed the main effect of training, $F(1, 124) = 15.49$, $p <. 01$, $d = 0.70$. Participants in the control group evaluated the instructor on this irrelevant dimension quite negatively ($M = 2.31$, $SD = 1.39$), whereas participants in the source monitoring training condition choose responses toward the middle point of the scale presumably because they could bring up neither favorable nor unfavorable experiences about the instructor on this irrelevant and inapplicable dimension ($M = 3.34$, $SD = 1.52$). Thus, ratings of the participants in the experimental group were closer to the ratings of the subject matter experts as measured through the distance measure (McIntyre, 1984) than the ratings of the participants in the control group; an ANOVA test on the distance measure confirmed the main effect of training, $F(1, 124) = 8.60$, $p < .01$. Thus, there was evidence suggesting that the control group participants might have relied on their overall impressions or their ratings on the relevant performance dimensions to a greater extent than the training group participants. Next, I examined the relevant correlations for a more conclusive answer on this question.

4.1.3.2. Correlations

If there is evidence of a greater halo effect in the control condition, then ratings of the instructor on the irrelevant dimension should correlate strongly with the ratings made along specific performance dimensions. In the source-monitoring training condition, this correlation should be significantly weaker than the correlation observed in the control condition. As can

be seen from the last rows of Tables 4.1 and 4.2, the ratings of the irrelevant dimension correlated significantly with all other specific and overall ratings in the control condition (minimum $r = .62$, maximum $r = .71$). It was not the case, however, in the source-monitoring training condition: The ratings of the irrelevant dimension did not correlate with any other ratings, (minimum $r = -.01$, maximum $r = .24$). Taken together, these data suggest that there was no evidence of halo effect in the training condition, but strong evidence in the control condition.

Table 4.1. Correlations among study variables in the control condition

	M	SD	1	2	3	4	5	6	7
Specific performance dimension ratings									
1. Ability to present material	2.54	1.46							
2. Interest in course and material	1.88	1.08	0.64**						
3. Relationships with students	2.37	1.30	0.69**	0.61**					
4. Fairness of testing and grading	2.48	1.60	0.45**	0.65**	0.68**				
5. Reasonableness of workload	2.17	1.17	0.63**	0.79**	0.73**	0.76**			
Overall ratings									
6. Initial overall evaluation	2.42	1.38	0.60**	0.69**	0.70**	0.68**	0.78**		
7. Delayed overall evaluation	2.26	1.28	0.56**	0.66**	0.63**	0.58**	0.74**	0.75**	
Irrelevant dimension ratings									
8. Ability to evaluate student progress	2.31	1.31	0.70**	0.64**	0.71**	0.66**	0.67**	0.65**	0.62**

Note: $N = 61$, $^{**}p < .01$.

Table 4.2. Correlations among study variables in the source-monitoring condition

			1	2	3	4	5	6	7
Specific performance dimension ratings	M	SD							
1. Ability to present material	3.12	1.39							
2. Interest in course and material	2.56	1.27	0.60**						
3. Relationships with students	2.92	1.15	0.42**	0.42**					
4. Fairness of testing and grading	2.88	1.10	0.44**	0.53**	0.46**				
5. Reasonableness of workload	2.88	1.22	0.56**	0.56**	0.56**	0.62**			
Overall ratings									
6. Initial overall evaluation	2.77	1.09	0.67**	0.63**	0.57**	0.49**	0.71**		
7. Delayed overall evaluation	3.02	1.40	0.43**	0.63**	0.38**	0.40**	0.52**	0.50**	
Irrelevant dimension ratings									
8. Ability to evaluate student progress	3.34	1.52	0.09	0.09	-0.01	0.24	0.18	0.15	0.12

Note: $N = 65$, $^{**}p < .01$

4.1.3.3. Reliabilities on the Performance Appraisal Form

The correlations presented above suggest that including the irrelevant dimension in the specific ratings scale would inflate the reliability coefficient for the control group, but decrease it for the training group. Table 4.3 displays the reliability coefficients on the performance appraisal form for both groups. As can be seen from the table, the reliability coefficient for the control group ($\alpha = .92$) was incomparably greater than the reliability coefficient for the training group ($\alpha = .77$). More importantly, an examination of the correlation of each dimension with the total score reveals that the source-monitoring training clearly decreases the magnitude of the halo error. For the control group, if the item measuring the ability to evaluate student progress and give feedback is deleted, the reliability coefficient does not change that much ($\alpha = .90$), whereas removing this item in the training condition increases the reliability coefficient considerably as it should ($\alpha = .84$).

Table 4.3. Reliability coefficients

	Item-Total Correlation (Control group)	Item-Total Correlation (Intervention group)
Specific dimensions		
Ability to present material	.71	.59
Interest in course and material	.78	.62
Relationships with students	.81	.50
Fairness of Testing and Grading	.73	.66
Reasonableness of workload	.84	.71
Irrelevant dimension		
Ability to evaluate student progress and give feedback	.79	.15
Overall reliability of the performance appraisal form	.92	.77

Table 4.3 also presents the item-total correlations for the scale. As can be seen, the item-total correlation of the irrelevant dimension in the control condition is incomparably greater than the correlation observed in the training condition (r = .79 vs. r = .15; z = 5.11, p <.001). In sum, an analysis of the reliability coefficients confirmed that there was strong evidence of halo effect only in the control condition.

4.1.3.4. Absolute Squared Deviations

Finally, halo can also be operationalized as the average of the absolute squared deviations of each dimension rating from the mean rating across dimensions (Balzer & Sulsky, 1992). When only the relevant items are taken into consideration, the average of the squared deviations of each dimension from the mean rating did not differ across the control and the training conditions. However, when the irrelevant item was taken into consideration, an ANOVA test revealed a significant difference between the two conditions, $F(1, 124)$ = 12.76, p <.001, d = 0.64. Thus, the ratings on the irrelevant dimension of the participants in control condition (M = 1.02, SD = 1.74) were significantly closer to the mean rating across dimensions compared to the participants in the source monitoring condition (M = 3.51, SD = 5.2). Overall, these findings support the first hypothesis that receiving source monitoring training would decrease the magnitude of halo error in a performance appraisal setting.

4.1.4. Recognition Memory

The judgment data reviewed above suggest that the training was useful in reducing the negativity bias and the associated halo effect. Thus, it is necessary to examine whether those

effects were indeed due to greater sensitivity to positive information in the training condition. Thus, participants in the control condition must have called a greater number of false alarms for negative behaviors and smaller number of hits for positive behaviors. In addition to hit rates and false alarm rates, memory strength was also computed.

4.1.4.1. The Effects of the Source Monitoring Training on Hit Rates and False Alarm Rates

The intervention effectiveness was expected to reflect on hit rates for positive behaviors (training > control) and false alarms for negative behaviors (control > training). Given that we expected to find negativity bias in the control group, finding a high hit rate for negative behaviors would not be surprising for this group; consequently, hit rates for negative behaviors were not diagnostic about the effectiveness of the intervention. Similarly, participants in this group would be conservative in finding something positive about the instructor, and hence false alarm rates for positive behaviors would be artificially low in the control group. Consequently, false alarm rates for positive behaviors were not diagnostic about the effectiveness of the interventions. Nonetheless, results for these data are also reported in the following section.

As expected, participants in the training group were more accurate in recalling the number of positive behaviors taking place in the video than the participants in the control group, $F(1, 124) = 4.38$, $p < .05$, $d = 0.40$. This difference in hit rates was not significant for negative behaviors as expected, $F(1, 124) = 1.90$, $p > .05$, $d = 0.26$. The data relevant to these analyses are reported in Table 4.4. As can be seen from the table, Hypothesis 4 was supported.

I expected that the participants in the control group would call a greater number of false alarms for negative behaviors than the participants in the training group. As expected, participants in the control condition reported more negative behaviors that did not occur than did participants in source monitoring condition, $F(1, 124) = 13.61$, $p<.001$, $d = 0.70$. This difference was not observed for positive behaviors $F(1, 124) = 1.35$, $p > .05$, $d = 0.22$. Taken together, these data support the fifth hypothesis that receiving source monitoring training would decrease the tendency to neglect positive attributes of a person having some negative attributes.

Table 4.4 Recognition memory measures across conditions

	Positive Behaviors					
	Hit rates		False-alarm rates		Memory strength	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control (n=61)	0.64	0.21	0.19	0.15	0.45	0.25
SMT (n=65)	0.70	0.14	0.16	0.13	0.55	0.19
Effect size (*d*)	0.34*		0.22		0.46**	

	Negative Behaviors					
	Hit rates		False-alarm rates		Memory strength	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control (n=61)	0.76	0.14	0.49	0.19	0.27	0.23
SMT (n=65)	0.79	0.10	0.36	0.18	0.43	0.21
Effect size (*d*)	0.26		0.70**		0.73**	

Note: Hit rate mean values range from 0 *(no "present" behaviors reported*) to 1.0 (*all "present" behaviors reported*). False alarm rates mean values range from 0 (*no "not present" behaviors reported*) to 1.0 (*all "not present" behaviors reported*). Memory strength values range from -1.0 (*no memory*) to 1.0 (*perfect memory*), * $p < .05$, ** $p < .01$

4.1.4.2. The Effects of the Source Monitoring Training on Behavioral Accuracy

The results on hit rates and false alarms point out the origins of bias in the control condition. In general, it is convenient to combine these two indices in assessing behavioral accuracy. Specifically, an index called memory strength is computed by subtracting false alarm rates from hit rates. Thus, memory strength represents a rater's ability to discriminate between behaviors that have been observed and that have not been observed (Snodgrass & Corwin, 1988). Using the memory strength for positive behaviors as the dependent variable, ANOVA was conducted. The main effect of intervention was significant, $F(1, 124) = 6.19$, $p < .05$, $d = 0.46$. The memory strength for positive behaviors of the participants in source monitoring condition ($M = 0.55$, $SD = 0.19$) was significantly higher than the memory strength for positive behaviors of the participants in control condition ($M = 0.45$, $SD = 0.25$).

As for negative behaviors, the main effect of intervention was again significant, $F(1, 124) = 16.21$, $p < .001$, $d = 0.73$. The memory strength for negative behaviors of the participants in the source monitoring condition ($M = 0.43$, $SD = 0.21$) was significantly stronger than the memory strength of the participants in control condition ($M = 0.27$, $SD = 0.23$). Overall, source monitoring training successfully increased the memory strength for both negative and positive behaviors.

As a final step, I combined false alarm rates and hit rates for positive and negative behaviors in order to measure overall behavioral accuracy. Thus, false alarms for positive and negative behavior were summed and an overall false alarm rate was calculated. Hits for both positive and negative behaviors were summed and an overall hit rate was calculated.

Behavioral accuracy was computed by subtracting false alarm rates from hit rates. The main effect of the training on this overall measure of behavioral accuracy was significant, $F(1, 124) = 14.78$, $p < .001$, $d = 0.68$. In conclusion, participants in the source monitoring training condition were better at distinguishing observed behaviors from unobserved behaviors than the participants in the control condition. Thus, Hypothesis 2 was supported.

4.2. Response Time Data

The recognition memory data summarized above suggest that participants in the source monitoring training condition must have searched their memory more deeply when appraising the performance of the instructor. Deeper search of memory for specific incidents can take longer than simply relying on overall impressions, expectations, and initial attitudes. Thus, participants in the training condition must have taken longer than participants in the control condition in reporting their appraisals of the instructor along specific performance dimensions. Analyzing the total amount of time taken in evaluating the instructor along specific performance dimensions revealed that participants in the training condition took significantly more time ($M = 67.68$ seconds, $SD = 31.27$) when evaluating the performance of the instructor than the participants in the control condition ($M = 43.33$ seconds, $SD = 17.65$), $F(1, 124) = 28.46$, $p < .01$, $d = 0.95$. This difference can be taken as evidence for the likelihood that participants in the training condition expanded greater effort in locating and sorting the relevant information.

4.3. Summary of Results

The present study examined the effectiveness of source monitoring training in reducing two well-known rater biases, namely the halo effect and the negativity bias. Overall, the results showed that source monitoring training can be a fruitful strategy in reducing the negativity bias and increasing the accuracy of performance appraisal decisions. The negativity bias was fueled by some of the same mechanisms guiding the halo effect. In that regard, by decreasing the extent of halo effect in appraisals, the training successfully brought about a decrease in negativity. In addition to judgment data, measures of recognition memory revealed that source monitoring training increased hit rates for positive behaviors and decreased false alarm rates for negative behaviors. In other words, source monitoring training enabled raters to discriminate between the behaviors that actually occurred from the behaviors that did not occur, and consequently increased rating and behavioral accuracy.

Chapter 5

DISCUSSION

The present study aimed to contribute to the performance appraisal training literature by examining the effectiveness of source monitoring training in decreasing primacy and halo effect. In the present study, it was expected that source monitoring training would decrease negativity bias. Furthermore, in the present study, it was expected that the ratings of the participants who received source monitoring training would show less halo than the ratings of the participants who did not receive any training. Additionally, the effect of source monitoring training in increasing overall accuracy and behavioral accuracy was examined. It was expected that participants who received the source monitoring training would produce ratings higher in rating and recognition accuracy than those who did not receive any training. Overall, these expectations were confirmed. Thus, source monitoring training appears to be a promising training method in decreasing rater biases.

5.1. Order of the Information

In this literature, when the performer displays both positive and negative behaviors, it is customary to control the order of these performances via counterbalancing. Part of the reason is that, there have been reports of both primacy and recency effects in past research (e.g., Kruglanski & Richter, 1998). In the present study, the order of presentation did not have

a significant effect on any of the dependent variables of interest. This finding can be explained as the difference of the type of the stimulus between this study and the other studies investigating order effects. In order to increase the generalizability of our results, a video was constructed specifically for this study. The scenes were arranged to reflect the reality. Furthermore, participants were instructed to watch the videotaped performance of the instructor closely and avoid making a premature judgment before watching the entire sequence. Earlier studies typically involved performance descriptions presented on paper, and without these instructions.

5.2. The Effect of Source Monitoring Training in Reducing Negativity Bias

In the present study, source monitoring training was found to reduce negativity bias, presumably because participants who received the training searched their memory for positive performances more deliberately. Consequently, they were able to remember positive behaviors besides negative behaviors and were more accurate in their decisions. This result can be interpreted in different ways.

First of all, with the help of behavioral questionnaire and the instructions asking the participants search their memory for remember judgments, source monitoring training probably increased the accessibility of not only negative but also the positive information. Smith et al. (2006) suggested that an intervention designed to decrease the relative accessibility of negative information in memory may reduce negativity bias. In line with this, as part of the training, participants were asked to search for their memory and rely on specific memory traces and not to use their know judgments. Consequently the training increased the

accessibility of the behavioral information, and hence not only behavioral accuracy but also the rating accuracy.

One possible explanation for observing negativity effects in the control condition is that negative information comes to mind more quickly than positive information. For instance, Ochsner (2000) found that negative stimuli were remembered more often compared to positive stimuli while positive stimuli were known more often compared to negative stimuli. This finding is relevant in the context of the present study. In the control condition, participants used their remember judgments that were mostly based on negative behavioral information. Participants who received the source monitoring training were asked to rely on specific memory traces rather than feelings of familiarity. Consequently, participants in the training condition were motivated to retrieve information in a relatively more even-handed way. Thus, they were able to avoid the biasing effects of negative behavioral information; as they retrieved and used both positive and negative pieces of information.

5.3. The Effect of Source Monitoring Training in Reducing Halo Effect

The present study also showed that the training can be useful in reducing or eliminating the halo effect as well. Across different measures of halo effect, the training was found to be effective. First of all, irrelevant dimension ratings of the participants who received the training were more accurate: The ratings in this condition were more neutral and accurate than the ratings in the control condition. In addition, the irrelevant dimension ratings of the participants who received the training instructions correlated strongly with neither the ratings on other specific and relevant ratings nor the overall impressions, but the ratings of the control group participants did. Furthermore, reliability coefficients confirmed that participants in the

control condition imputed their overall ratings to irrelevant dimension whereas the participants in the training condition did not impute their overall ratings to the rating of irrelevant dimension. Finally, by using another halo measure, I found that the ratings based on the irrelevant dimension of the participants in control condition were significantly closer to their mean ratings across dimensions compared to the participants in source monitoring condition. In sum, the training was useful in reducing the halo effect.

5.4. Recognition Memory Improvement with Source Monitoring Training

In the present study, various recognition memory indices (i.e., hit rates, false alarms, memory strength) were computed to assess how well participants in each conditions retrieved specific memory traces while evaluating the instructor. Results showed that hit rates for positive behaviors were significantly greater in the training condition, whereas false alarms for negative behaviors were greater in the control condition. Thus, the source monitoring training can be said to make participants more sensitive to positive information. Instead of assuming that the instructor was a poor performer overall like control group participants, these participants searched their memory for positive incidents and adjusted their ratings accordingly; their ratings were relatively accurate compared to the ratings of control group participants. The effect of the training on negative hit rates was not as pronounced as its effect on positive behaviors. The baseline hit rates for negative behaviors were already so high; presumably there was not much room for improvement for negative behaviors. As noted, control group participants developed a negative impression of the instructor and relied on this impression in filling out the forms. Consequently, they called significantly more false alarms than the training group participants.

This difference in positive and negative hit rates can be attributed to the fact that negative information is typically weighted more heavily than positive (Baumeister et. al., 2001). Besides, being "cognitive misers", people are not able to process all of the information at the same extent; thus they must determine the top-priority information sources and focus on the most important aspect. Consequently, the participants possibly thought that the instructor performance was not sufficient enough. Thus, since the role of an instructor is very important for them, the participants processed specifically negative information. The hit rates for negative information were very high and could not be increased further more by source monitoring training instructions.

Finally, source monitoring training increased the memory strength for positive and negative behaviors. In order to measure overall behavioral accuracy, false alarm rates and hit rates for positive and negative behaviors were combined. The results on this overall measure of recognition accuracy clearly showed that source monitoring training was effective in increasing accuracy. Participants in the training condition were better at distinguishing observed behaviors from unobserved behaviors compared to participants in the control condition. This finding is very crucial. Performance appraisal training research suggested that a primary purpose of much performance appraisal is to provide specific feedback for training and development of the employees, consequently behavioral accuracy is very critical (Woehr & Huffcutt, 1994). Therefore, by making raters rely on remember judgments based on specific memory traces while ignoring know judgments based on feelings of familiarity (Martell & Evans, 2005), source monitoring training appears to be a very useful and effective tool for performance appraisals in organizations.

5.5. Contributions of the Present Study

There are various training programs designed to increase the accuracy of performance appraisal decisions. Some of the programs, however, are not practical enough for use in field settings (for a review of such interventions, see e.g., Woehr & Huffcutt, 1994). There is need for practical and effective training programs. In that regard, the source monitoring training appears to be a promising one that practitioners may found to be appealing, not only because of its effects but also because of its ease of implementation. It is a very effective and practical method and also it is not time consuming. The present study showed that it can be used in a performance appraisal setting. As mentioned, the negativity bias and the halo effect are quite widely observed in everyday settings. Therefore, these findings will be of interest to practitioners interested in improving their performance appraisal systems.

5.6. Limitations of the Present Study and Suggestions for Future Research

The present study has a number of limitations. For instance, the present study involved student participants. Although, the performances evaluated by the students were as realistic as possible, students may still be different from managers in work settings. Lecturers work for teaching the students while employees report to the managers. Given the findings of the present study, I believe that it would be worthwhile to examine the effects of the source monitoring training in a field study.

Another limitation of the present study was related to using two types of video: The first one begins with positive behaviors followed by negative behaviors whereas the second one begins with negative behaviors followed by positive behaviors. Even though the results

did not indicate a significant order effect, the chunk of negative behaviors could have artificially increased the extent of negativity bias. In other words, by observing a block of negative behaviors, participants could have developed an irreversible and strong negative impression towards the instructor. Therefore, using a mixed video and comparing it with the other two videos can be useful. Thus, future research can examine the effects of source monitoring training by using an additional condition in which the participants observed a video composed of randomly selected instructor behaviors.

The present study has another limitation. As in most other studies in the literature, participants in the present study rated the performance of a single person. In real world settings, students completing teaching evaluations and manager completing end-of-year performance evaluation forms have to do that for more than one ratee. Thus, it is likely that contrast effects take place in everyday life because of comparative evaluations. Contrast effect occurs when raters rate the performance away from the direction of the ratee's previous performance (Sümer & Knight, 1996). The effectiveness of any intervention should be examined in such multiple-ratee settings; unfortunately, the present study shares this limitation with most of the other studies in the literature.

As a future direction, it is also necessary to compare the effectiveness of the source monitoring training with other types of trainings or interventions used in performance appraisal settings. For instance, source monitoring training should be compared with the frame-of-reference training. This point was also indicated by Martell and Evans (2005) that in order to determine the practical benefit of source monitoring training, one should compare it with other trainings or interventions.

In the current study, participants rated the performance of the ratee once and solely at the end of the performance. Today, companies may ask raters conduct performance appraisals twice a year. Thus, managers observe the ratees, conduct interim reviews, continue to observe the performances and give final ratings at the end of the year. Managers conducting appraisals twice may make different decisions compared to a situation that they make ratings solely at the end of the year. For example, Sümer and Knight (1996) found evidence of both assimilation and contrasts effects. When raters did not evaluate the previous performance of the ratee, they were likely to assimilate the subsequent performance because much of what they recalled was congruent with the previous impression. When they rated the previous performance of the ratee, however, there was evidence of a contrast effect. Thus, the participants who have rated the prior performance first, rated the performance away from the prior performance at the end. Thus, further examination of the effects of the source-monitoring training when raters make performance appraisals more than once is worthwhile.

Lastly, in the present study there was not a condition in which source monitoring training was combined with another training approach. There is need for further research investigating the effects of source monitoring training combined with other methods. For instance, source monitoring training can be combined with frame-of-reference training. The study of Sulsky and Day (1992) showed that frame-of-reference training increases rating accuracy while decreasing behavioral accuracy. Furthermore, frame-of-reference training does not increase the quality of the behaviors that raters recalled (Roch & O'Sullivan, 2003). Decreased behavioral accuracy and the quality of the behaviors recalled decrease the accuracy of performance feedback. On the other hand, perceptions of justice have been linked to many important organizational variables such as organizational commitment (Tang & Sarsfield-

Baldwin, 1996) and organizational citizenship behavior (Niehoff & Moorman, 1993). Thus, increased behavioral accuracy and increased quality of feedback is very important of organizations. Thus, the accuracy can be further increased by frame-of-reference training. Also source monitoring training can be used to increase behavioral accuracy.

It is also likely that the source monitoring training may be supported by diary keeping method. Source monitoring training was introduced just after the observational stage of behavioral measurement process in Martell and Evans (2005)'s study. However, depending on the available information from activated memory records, source monitoring relies vitally on the quality of the recorded information about events (Johnson, Hastroudi, & Lindsay, 1993). It is very important that during performance appraisal process, managers have information recorded well in order to make correct source attributions. Therefore, it is clear that raters also need some tools not only during rating but also during observational stage. In addition, if a memory response is experienced sufficiently and uniquely specifying a moment in one's past, it is experienced as a remember judgment (Bodner & Lindsay, 2003). Authors suggested that the possibility that people will remember something from their past may depend in part, on how fine they remember the answers to the other questions in the interview in any multi-question interview situation. If managers have some notes indicating past behaviors of the ratee in front of them, it can be easier for them to remember other behavioral incidences even if these incidences may not be in the diary.

Source monitoring can rely not only on recollection but also on qualitative characteristics that lack clarity and sufficient amounts of details in order to increase the subjective feeling of remembering (Hicks et al., 2002). In that regard, because source monitoring decisions are likely to be based on mixture of recollective information and some

additional vague details (Hicks, Marsh, & Ritschel, 2002), it may be possible to obtain more accurate source monitoring decisions by increasing the amount of information and level of clarity. If diary keeping is used properly by the raters it may be useful to combine it with source monitoring training as it facilitates the encoding of performance related information. In conclusion, the combination of these two methods can produce the most effective results in performance appraisal settings. Future research examining the effectiveness of such combined interventions would be definitely worthwhile.

In conclusion, the present study showed that source monitoring training works well in reducing two important biases that take place at the time of making a judgment. The difference between the ratings of the participants in the training and control conditions cannot be attributed to differential encoding because the intervention was introduced after encoding, but to differential retrieval and selection of inputs at the time of making a judgment. In that regard, the source-monitoring training can be useful in reducing other errors and biases that take place at the time of making a judgment. For instance, there is evidence suggesting that people can use their affective states (e.g., mood) as information in making a judgment, and rate a person more favorably if they are in a happy mood or more unfavorably if they are in a negative mood (Albarracín & Kumkale, 2003). It would be worthwhile to examine the effectiveness of source-monitoring training in reducing such biases as well.

REFERENCES

Albarracín, D., & Kumkale, G. T. (2003). Affect as information in persuasion: A model of affect identification and discounting. *Journal of Personality and Social Psychology, 84*, 453-469.

Ambady, N., & Gray, H. M. (2002) On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83,* 947-961.

Anderson, C. A., & Slusher**,** M. P. (1987). When reality monitoring fails: The role of imagination in stereotype maintenance. *Journal of Personality and Social Psychology, 52,* 653-662.

Baltes, B. B., & Parker, C. P. (2000). Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior & Human Decision Processes*, *82*, 237–267.

Baltes, B. B., & Parker, C. P. (2000) Understanding and removing the effects of performance cues on behavioral ratings. *Journal of Business and Psychology, 15(2),* 229-246.

Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology, 77,* 975-985.

Bauer, C. C., & Baltes, B. B. (2002). Reducing the effects of gender stereotypes on performance evaluations. *Sex Roles, 47*, 465-476.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5(4),* 323-370.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory & Language*, *48*, 563–580.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology*, *63*, 135–144.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119,* 197-253.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2003). Halo and devil effects demonstrate valanced-based influences on source-monitoring decisions. *Consciousness and Cognition, 12*, 257-278.

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance*, *33*, 457–467.

DeNisi, A. S. & Peters, L. H. (1996). Organization of information in memory and the performance appraisal process: evidence from the field. *Journal of Applied Psychology, 81*, 717-737.

Dessler, C. (2005). Human Resource Management. New Jersey: Pearson Education International.

Feldman, J. M. (1981). Beyond attribution theory: cognitive processes in performance appraisal. *Journal of Applied Psychology*, *66,* 127-148.

Feldman, J. M. (1986). Instrumentation and training for performance appraisal: A perceptual–cognitive viewpoint. In K.Rowland & G.Ferris (Eds.), *Research in personnel and human resources management* (*Vol. 4*, pp. 45–100). Greenwich, CT: JAI Press.

Feldman, J. M. & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention and behavior. *Journal of Applied Psychology, 73*, 421-435.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*, 889-906.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309-313.

Gardiner, J. M., & Java, R. I. (1993). Recognition memory and awareness: An experiential approach. *European Journal of Cognitive Psychology*, *5*, 337–346.

Goffin, R. D., Jelley, R. B., & Wagner, S. H.. (2003) Is halo helpful? Effects of inducıng halo on performance ratıng accuracy. *Social Behavior & Personality, 31,* 625-637.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4,* 106-112.

Haugtvedt, C. P., & Wegener, D. T. (1994). Message order effects in pesuasion: An attitude strength perspective. *Journal of Consumer Research, 21,* 205-218.

Hekkanen, S. T., & McEvoy, C. (2002) False memories and source-monitoring problems: criterion differences. *Applied Cognitive Psychology, 16,* 73-85.

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 503-508.

Ikegami, T. (1993). Positive-negative asymmetry of priming effects on impression formation. *European Journal of Social Psychology, 23,* 1-16.

Ito, T.A., & Cacioppo, J. T. (2005). Variations on a human universal: Individual differences in positivity ofset and negativity bias. *Cognition and Emotion, 19*, 1-26.

Jelley, R.B., & Goffin, R. D. (2001). Can performance-feedback accuracy be improved? Effects of rater priming and rating-scale format on rating accuracy. *Journal of Applied Psychology, 86*, 134-144.

Johnson, J. T., & Judd, C. M. (1983). Overlooking the incongruent: Categorization biases in the identification of political statements. *Journal of Personality and Social Psychology, 45(5),* 978-996.

Johnson, M. K., Hastroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Review*, *114*, 3–28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88,* 67-85.

Kahan, T. L., & Johnson, M. K. (1992). Self effects in memory for person information. *Social Cognition, 10,* 30-50.

Lance, C. E., LaPointe, J.A., & Stewart, A. M. (1994). A test of the context dependency of three causal models of halo rater error. *Journal of Applied Psychology*, *79*, 332-340.

Lord, R. G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology*, *70*, 66–71.

Manzoni, J. F., & Barsoux, J. L. (1998). How bosses create their own poor performers: The set-up-to-fail syndrome. *Harvard Business Review,* 101–113.

Martell, R. F., & Evans, D. P. (2005). Source-monitoring training: Toward reducing rater expectancy effects in behavioral measurement. *Journal of Applied Psychology, 90,* 956-963.

Martell, R. F., Guzzo, R. A., & Willis, C. E. (1995). A methodological and substantive note on the performance-cue effect in ratings of work group behavior. *Journal of Applied Psychology*, *80*, 191–195.

Martell, R. F., & Willis, C. E. (1993). Effects of observers' performance expectations on behavior ratings of work groups: Memory or response bias?*Organizational Behavior & Human Decision Processes*, *56*, 91–109.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology, 69,* 147-156.

Mirels, H. L., Greblo, P., & Dean, J. B. (2002). Judgmental self-doubt: Beliefs about one's judgmental prowess. *Personality and Individual Differences, 33*, 741-758.

Murphy, K. R. & Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: consequences for rating accuracy. *Journal of Applied Psychology, 71*, 39-44.

Murphy, K. R. & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, goal-based perspectives*. London: Sage Publications.

Murphy, K.R., & Jako, R. (1989). Under what conditions are observed intercorrelations greater or smaller than true intercorrelations? *Journal of Applied Psychology, 74*, 827-830.

Murphy, M. R., & Jones, A. (1993). The influences of performance cues and observational focus on performance rating accuracy. *Journal of Applied Social Psychology, 23,* 1523–1545.

Niehoff, B.P. & Moorman, R.H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal, 36*, 527-556.

Noonan**,** L. E., & Sulsky**,** L. M. (2001). Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Human Performance, 14*, 3-26

Oden, G.C., & Anderson, N.H. (1971). Differential weighting in integration theory. *Journal of Experimental Psychology*, *89,* 152-161.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129,* 242-261.

Palmer**,** J., Thomas, A., & Maurer, T. (2003). Moderating effects of context on the relationship between behavioral diaries and performance rating halo and accuracy. *North American Journal of Psychology, 5,* 81-90.

Pratto, F. & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-391.

Richter, L. & Kruglanski, A. W. (1998). Seizing on the latest: motivationally driven recency effects in impression formation. *Journal of Experimental Social Psychology, 34,* 313-329.

Roch, S. G., & O'Sullivan, B. J. (2003). Frame of reference rater training issues: recall, time and behavior observation training. *International Journal of Training and Development, 7*, 93-107.

Sanchez J.I. & De la Torre, P. (1996). A second look at the relationship between rating and behavioral accuracy in performance appraisal. *Journal of Applied Psychology*, *81*, 3-10.

Sauser, W. I., Jr., Evans, K. L., & Champion, l. C. H. (1979). Two hundred and fifty scaled incidents of college classroom teaching behavior. *Resources in Education*, 15(1), 30.

Sherman**,** J. W., & Bessenoff, G. R. (1999) Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science, 10,* 106-111.

Skowronski, J.J., & Carlston, D.E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*, 131-142.

Smith, N.K., Larsen, J.T., Chartrand, T.L., Cacioppo, J.T., Katafiasz, H. A., & Moran, K. E. (2006). Being bad isn't always good: Affective content moderates the attention bias toward negative information. *Journal of Personality and Social Psychology, 90,* 210-220.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia*. Journal of Experimental Psychology: General, 117*, 34-50.

Srull, T. K., & Wyer, R. S. (1983) The role of control processes ans structural constraints in models of memory and social judgment. *Journal of Experimental Social Psychology, 19,* 497-521.

Sulsky, L. M., & Day, D. V. (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology*, *77*, 501–510.

Sulsky, L. M.,&Day, D. V. (1994). Effects of frame-of-reference training on rater accuracy under alternative time delays. *Journal of Applied Psychology, 79,* 535–543.

Sümer, H. C., & Knight, P. A. (1996). Assimilation and contrast effects in performance ratings: Effects of rating the previous performance on rating subsequent performance. *Journal of Applied Psychology*, *81*, 436–442.

Tang, T.L. & Sarsfield-Baldwin, L. J. (1996). Distributive and procedural justice as related to satisfaction and commitment. *S.A.M. Advanced Management Journal, 61*, 25-32.

Thierry, K. L.& Spence, M. J. (2002) Source-monitoring training facilitates preschoolers' eyewitness memory performance. *Developmental Psychology, 38*, 428-437.

Thorndike, E. L. (1920). A constant error on psychological rating. *Journal of Applied Psychology, 4*, 25-29.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1–12.

Webster, D. M. & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*, 1049-1062.

White, J.D., & Carlston, D. E. (1983). Consequences of schemeta for attention, impressions and recall in complex social interactions. *Journal of Personality and Social Psychology, 45,* 538-549.

Woehr, D. J., & Feldman, J. (1993). Processing objective and question order effects on the causal relation between memory and judgment in performance appraisal: Tip of the iceberg. *Journal of Applied Psychology, 78,* 232–241.

Woehr, D. J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of Applied Psychology, 79*, 525–534.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.

Wojciszke, B.; Brycz, H.; Borkenau, P. (1993). Effects of information content evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology, 64*, 327-335

APPENDICES

APPENDIX A

Instructions Given to Participants

"Öncelikle deneye katıldığınız için çok teşekkür ederiz Bu çalışma üniversitelerarası bir projenin parçasıdır. Farklı üniversitelerden öğrencilerden toplanan datalarla farklı öğretim üyelerinin performansı değerlendirilmektedir. Birazdan izleyeceğiniz videodaki öğretim üyesi bir yıl boyunca değişik zamanlarda gizlice kameraya alınmış ve performansını en iyi bir şekilde gösterecek video kesitleri size sunulmuştur. Bu çalışma sırasında sizden beklenen videoyu iyice izleyip sonrasında da öğretim üyesinin performansını değerlendirmektir. Lütfen videoyu izleyin ve bu süreçte arkadaşlarınızla konuşmayın, herhangi bir not almayın.

Video başlatılır.

Aralarda durdurulur ve katılımcılara google logoları gösterilir ve bu logolara karşı olan beğenilerini bir ölçek üzerinde belirtmeleri beklenir. Bu aşamadan sonra bir matematiksel işlem verilir ve bu işlemin sonucunun kendilerine sorulacak diğer işleme kadar akıllarında tutmaları istenir.

Video biter. Deney grubundaki katılımcılara eğitim yönergeleri sunulur. Kontrol grubundaki katılımcılar ise aradaki süreyi eşitlemek için ölçek değerlendirmesine alınırlar.

Eğitim Yönergeleri

Bazen bir şeyi tanıdığımızı düşünürüz çünkü o şeyi aklımıza getirip zihnimizde canlandırabiliriz. Mesela geçen hafta bir partide tanıştığınız bir insanı sokakta gördüğünüzü

hayal edin. Bu insanla tanıştığımıza dair zihninizde bir görüntü oluştuğunun farkına varırsınız. Bu görüntü, o kişiyle olan konuşmanızın anısını, o kişiyle karşılaştığınızda neler düşünüyor olduğunuzu veya partide o kişiyle beraber nerede durduğunuzu içerebilir. Bu bir hatırlama yargısıdır; bu yargıyla geçmişte gerçekleşmiş bir olayı zihnimizde yeniden yaşarız.

Hatırlanan bir olayı aklımıza belirli bir çağrışımı, görüntüyü, veya o olayla ilgili başka şeyleri aklımıza getirmemizi sağlar. Böylece o olayı hatırlama durumu gerçekleşir.

Buna rağmen başka zamanlarda, bir şeyi, sırf bize aşina geldiği için, kendimize güvenerek hatırladığımızı düşünürüz. Böylesi durumlarda, tam olarak olayı aklımıza getirip, zihnimizde canlandıramasak da, olay bize aşina geldiğinden olayın gerçekleşmiş olduğuna eminizdir. Bu bir bilme yargısıdır. Örneğin, yine geçen hafta bir partide tanıştığınız biriyle sokakta karşılaştığınızı düşünün. Fakat bu sefer aslında o kişiyle partide tanıştığınızı zihninizde canlandıramıyorsunuz. Buna rağmen o insanı partide tanıştığınız insan olarak tanıdığınıza emin olabilirsiniz çünkü o insanı görmeniz aşinalık duygunuzu tetikler. Bu bir bilme yargısıdır. Zihnimizde bir olayı yeniden yaşamış gibi olamadığımız zamanlarda bile olayın gerçekleşmiş olduğundan eminsek bu tür bir yargıya sahibiz demektir. Çoğu insan için zihninde bir olayı canlandıramadan insanları, yerleri, nesneleri ve davranışları tanımak sıkça rastlanan bir durumdur.

Şunu anlamak çok önemlidir; bir bilme yargısına sahip olduğumuzda, hatırlama yargısına sahipmiş gibi kendimizden emin olabiliriz. Örneğin, iki olayın, diyelim ki A ve B olayının gerçekleşmiş olduğu konusunda eşit derecede emin olabiliriz. A olayından emin olmamız A olayını hatırlamamıza, B olayını hatırlamamız B olayının gerçekleştiğini bilmemize dayanıyor olsa bile iki olaya ait belleğimizden eşit derecede emin olabiliriz.

Bir hatırlama yargısına sahip olmanız durumunda videoda izlemiş olduğunuz öğretim elemanının davranışlarını bilinçli olarak hatırlayabilirsiniz ve bu davranışları zihninizde canlandırabilirsiniz. Bu tür bir yargı bu davranışın gösteriliş şekline dair hatırladıklarınızı, bu davranışı gözlemlerken ne düşünüyor olduğunuzu veya davranışı gördüğünüzde ne hissettiğinizi içerebilir. Hatırlanan bir davranış, aklınıza belirli bir çağrışımı, görüntüyü, sesi veya öğretim elemanının davranışıyla alakalı bir görüntüyü getirmelidir.

Eğer bir davranışı görmüş olmanızla ilgili spesifik ayrıntılar hatırlayabiliyorsanız, o davranışı daha önce görmüş olmanızın uyandırdığı düşünceleri ve duyguları aklınıza getirebilirsiniz.

Bir bilme yargısına sahipseniz videoda izlemiş olduğunuz öğretim elemanının davranışını davranışa olan aşinalık duygunuzdan dolayı tanıyorsunuzdur. Bu tür durumlarda, öğretim üyesinin davranışını videoda gözlemlemiş olduğunuza dair bir tecrübenizi ayrıntılarıyla hatırlayamıyor olsanız bile, o davranışa aşina olmanızdan dolayı onu videoda görmüş olduğunuzdan eminsinizdir.

APPENDIX B

Performance Evaluation Form

Az önce bir öğretim elemanının çeşitli davranışlarını içeren bir video izlediniz. Aşağıdaki 7 basamaklı ölçeği kullanarak bu öğretim elemanının performansını değerlendiriniz.

1	2	3	4	5	6	7
Çok zayıf			**Orta**			**Mükemmel**

1. Bu öğretim elemanının genel performansı...

2. Öğretim elemanının öğrencilerine destek olma ve danışmanlık etme davranışı...

3. Öğretim elemanının yardım istendiğinde ulaşılabilir olması..

4. Öğretim elemanının ders anlatımının organizasyonu...

5. Dersin açıklığı ve takip edilebilirliği...

6. Öğretim elemanının derse karşı ilgisi ve isteği...

7. Öğretmenin dersin içeriğine hakimiyeti...

8. Öğretim elemanının ders yükünü dengeleyebilme ve dağıtabilme becerisi...

9. Öğretim elemanının yapılması gereken işin niteliğini belirtme ve işin zamanlamasını ayarlama becerisi...

10. Öğretim elemanının sınav formatını ve soruları hazırlamadaki adaleti...

11. Öğretim elemanının sınavları değerlendirmedeki adaleti...

1	2	3	4	5	6	7
Kesinlikle katılmıyorum			**Ne katılıyorum Ne Katılmıyorum**			**Kesinlikle katılıyorum**

12. Öğretmenin öğrencilerin performansını takip edebilme becerisi iyidir.

13. Öğretmenin öğrencilerin performansı hakkında onlara geribildirim verebilme becerisi iyidir. (feedback)

APPENDIX C

Behavioral Questionnaire

Birazdan size videonun içeriği ile ilgili sorular soracağız. Size bir takım davranışlar sıralayacağız ve bu davranışların videoda gerçekleşip gerçekleşmediğine karar vermenizi istiyoruz.

Örneğin, "sınıfta kavga çıktı ve öğretim elemanı bunu durdurmadı" cümlesini size vererek öğretim elemanının bu davranışının gerçekleşip gerçekleşmediğini size soruyor olacağız.

Bu ölçek sizin bu davranışın gerçekleşip gerçekleşmediğine ne derece emin olduğunuzu ölçer.

Bir davranışın gerçekleşmediğine çok eminseniz, 1'i işaretleyin. Gerçekleşmediğini düşünüyorsanız ve eminseniz 2'yi işaretleyin. Emin değilseniz ama bir şekilde gerçekleşmediğini düşünüyorsanız 3'ü işaretleyin. Emin değilseniz ama gerçekleştiğini düşünüyorsanız 4'ü işaretleyin. Gerçekleştiğine eminseniz, 5'i işaretleyin. Davranışın gerçekleştiğine çok eminseniz 6'yı işaretleyin.

Şimdi davranışları okumaya geçebilirsiniz. Lütfen devam (continue) basarak devam edin.

6=çok eminim oldu / 1=çok eminim olmadı

	1	**2**	**3**	**4**	**5**	**6**
Bu öğretim elemanı sınıfta anlattığı şeylere şahit olmuştu ve kişisel tecrübelerini sınıf ile paylaştı.	1	2	3	4	5	6
Bu öğretim elemanı sınıfa okuduğu ya da duyduğu ilginç makalelerden ve deneylerden bahsetti.	1	2	3	4	5	6
Bu öğretim elemanı sınavlar arasında iki hafta boşluk bıraktı.	1	2	3	4	5	6
Bu öğretim elemanı ofis saatlerini öğrencilere duyurdu.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilere soru sorarak aldıkları diğer dersler ile kendi dersi arasında bağlantı kurmalarını sağladı.	1	2	3	4	5	6
Bu öğretim elemanı diğer fakülte üyeleriyle iletişime geçerek öğrencilerin sorunlarını çözmeye çalıştı.	1	2	3	4	5	6
Bu öğretim elemanı sınav döneminde öğrencilerin yükünü hafifletmek için verdiği ödevlerin bazılarını iptal etti.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilerin kendisine aktardığı fikirlerden dolayı onları takdir etti.	1	2	3	4	5	6
Bu öğretim elemanı derste anlattıklarını sınavda daha iyi hatırlamaları için öğrencilere taktik verdi.	1	2	3	4	5	6
Bu öğretim elemanı önceki derste ne işlediğini özetleyerek yeni konuya geçti.	1	2	3	4	5	6
Bu öğretim elemanı sömestr sonuna doğru ekstra bir ödev verdi.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilerle sınıf dışında konuşmayı çok dolu olan programı yüzünden reddetti.	1	2	3	4	5	6
Bu öğretim elemanı dersle ilgili sorular sorduklarında öğrencileri azarladı.	1	2	3	4	5	6
Bu öğretim elemanı sınavda soracağım dediği yerden değil de başka bir yerden soru sordu.	1	2	3	4	5	6
Bu öğretim elemanı yalnızca para kazanmak için öğretmenlik yaptığını söyledi.	1	2	3	4	5	6
Bu öğretim elemanı kendisinden talep edildiği halde öğrencilerin kağıtlarına tekrar bakmayı reddetti.	1	2	3	4	5	6
Bu öğretim elemanı yetiştiremediği konulardan da öğrencileri sınavda sorumlu tuttu.	1	2	3	4	5	6
Bu öğretim elemanı ders ile ilgili hata yaptıklarında öğrencileri azarladı.	1	2	3	4	5	6
Bu öğretim elamanı ofis saati için haftanın sadece bir saatini ayırdı.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilerden ders ile ilgili gelen farklı fikirleri sert bir şekilde eleştirdi.	1	2	3	4	5	6
Bu öğretim elemanı dört beş sayfalık bir ödev verdi ve öğrencilere bu ödevin nasıl yapılacağına dair bilgi sağladı.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilere, öğrencilerin kariyerleri ile ilgili yol gösterdi.	1	2	3	4	5	6
Bu öğretim elemanı sınıfta güncel konulardan birini öğrenciler ile tartışmak istedi.	1	2	3	4	5	6
Bu öğretim elemanı ders anlatırken öğrencilerin dikkatini çekebilecek çeşitli yüz ve vücut hareketleri kullandı.	1	2	3	4	5	6
Bu öğretim elemanı sınavda çeşitli soru formatları kullanacağını belirtti.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilerin ismini öğrenmeyi denedi.	1	2	3	4	5	6

Bu öğretim elemanı sınıfın isteği doğrultusunda verdiği ödevi bir gün erteledi.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilerin konuyu anladıklarından emin olmak için anlayıp anlamadıklarını sordu.	1	2	3	4	5	6
Bu öğretim elemanı ders anlatırken çeşitli yardımcı şekiller kullandı.	1	2	3	4	5	6
Bu öğretim elemanı haftanın başında o haftanın ödevlerini verdi.	1	2	3	4	5	6
Bu öğretim elemanı öğrencilerden iki üniteyi birden bir gecede okumalarını istedi.	1	2	3	4	5	6
Bu öğretim elemanı güncel gelişmeleri takip etse de bunları sınıfla paylaşmadı.	1	2	3	4	5	6
Bu öğretim elemanı geçen ders nerede kaldığını hatırlayamadı.	1	2	3	4	5	6
Bu öğretim elemanı verdiği ödevlerden birinin teslim tarihini bir hafta öne aldı.	1	2	3	4	5	6
Bu öğretim elemanı iki öğrencinin cevapları benzer olsa da kısa yazan öğrenciye uzun yazan öğrenci ile aynı puanı vermeyi reddetti.	1	2	3	4	5	6
Bu öğretim elemanı laboratuarda önceden işlenmiş bir konuyu yine de öğrencilere anlattı.	1	2	3	4	5	6
Bu öğretim elemanı ders sırasında öğrencilerden gelen bir soruya cevap vermeyi erteledi.	1	2	3	4	5	6
Bu öğretim elemanı ofisinde arandığı zaman bulunamadı.	1	2	3	4	5	6
Bu öğretim elemanı sınıfın kendisine açıkladığı bir sorunla ilgili herhangi bir çalışma yapmadı.	1	2	3	4	5	6
Bu öğretim elemanı kitapta olmayan bir konu hakkında öğrenciden gelen soruya cevap veremedi.	1	2	3	4	5	6

Şimdi son olarak tekrar geriye dönüp düşünmenizi istiyoruz. Öğretmenin performansını size sorulan sorular ile değerlendiriniz. Lütfen performans değerlendirmelerini size verilecek ölçeklere göre yapınız.

Bu öğretim elemanının genel performansı...

1	**2**	**3**	**4**	**5**	**6**	**7**
Çok zayıf			**Orta**			**Mükemmel**