

12-2017

# Putting Your Best Face Forward: The Influence of Facial Cosmetics on Structured Employment Interview Ratings

Skye K. Gillispie  
*Clemson University*

Follow this and additional works at: [https://tigerprints.clemson.edu/all\\_dissertations](https://tigerprints.clemson.edu/all_dissertations)

---

## Recommended Citation

Gillispie, Skye K., "Putting Your Best Face Forward: The Influence of Facial Cosmetics on Structured Employment Interview Ratings" (2017). *All Dissertations*. 2066.

[https://tigerprints.clemson.edu/all\\_dissertations/2066](https://tigerprints.clemson.edu/all_dissertations/2066)

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact [kokeefe@clemson.edu](mailto:kokeefe@clemson.edu).

PUTTING YOUR BEST FACE FORWARD: THE INFLUENCE OF FACIAL  
COSMETICS ON STRUCTURED EMPLOYMENT INTERVIEW RATINGS

---

A Dissertation  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy  
Industrial-Organizational Psychology

---

by  
Skye K. Gillispie  
December 2017

---

Accepted by:  
Dr. Robert Sinclair, Committee Chair  
Dr. Patrick Raymark  
Dr. Thomas Britt  
Dr. Mary Anne Taylor

## ABSTRACT

Employment interviews are ubiquitous in modern selection systems. Although interviews are extremely common, there is evidence that interview ratings are subject to rating errors and biases. For example, previous research has found that higher physical attractiveness of the candidate is linked to increased interview ratings. Physical attractiveness is largely considered to be a fixed characteristic that cannot be controlled, however this may not be entirely true as research has consistently linked women's use of facial cosmetics to increased ratings of physical attractiveness. An experimental three (no cosmetics, low cosmetics, high cosmetics) by three (low performance, intermediate performance, high performance) design was used to examine: a) what amount of facial cosmetics is most beneficial to interview ratings, b) the explanatory mediators of the cosmetics-interview ratings relationship, and c) the influence of interview performance on the cosmetics-interview ratings relationship. Participants included 452 individuals recruited using Amazon's Mechanical Turk. Results indicated that there was not a direct relationship between facial cosmetics use and interview ratings, but facial cosmetics did indirectly affect interview ratings through the mediating variables of physical attractiveness and professional appearance. Ratings of professional appearance were highest in the low cosmetics condition, suggesting that the amount of makeup worn affects perceptions of professional appearance. Contrary to expectations, facial cosmetics did not affect perceived competence, perceived competence did not mediate the relationship between facial cosmetics and interview ratings, and interview performance did not moderate the relationship between facial cosmetics and interview ratings. Overall,

the results of this dissertation provide some support for the common advice that it is important for women to wear makeup to job interviews.

## ACKNOWLEDGEMENTS

I would first like to thank my adviser, Dr. Bob Sinclair, for his thoughtful feedback and patience as I completed this dissertation, as well as my committee members, Drs. Patrick Raymark, Thomas Britt, and Mary Anne Taylor. You have each taught me so much throughout my graduate career, and I am so thankful for the education you have helped to provide me. I would also like to thank the Clemson University Graduate School and Psychology Department for funding this research. Additionally, I must thank the Cuyahoga County Personnel Review Commission for supporting me in my continued efforts to pursue my doctoral degree. Lastly, I want to thank the many friends and family members that offered support and encouragement as I completed this dissertation. Specifically, I owe a debt of gratitude to my husband, Andrew. I am so grateful for all of your love and support.

## TABLE OF CONTENTS

	Page
TITLE PAGE .....	i
ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTER	
I. INTRODUCTION .....	1
Study Contributions .....	2
Interview Psychometric Properties .....	5
Interview Ratings .....	15
Appearance .....	21
Interaction between Performance and Cosmetics .....	32
The Current Study .....	33
II. PILOT STUDY ONE .....	36
Method .....	36
Results .....	39
Discussion .....	40
III. PILOT STUDY TWO .....	41
Method .....	41
Results .....	45
Discussion .....	48
IV. STUDY THREE .....	49
Method .....	49
Results .....	54
Discussion .....	61

Table of Contents (Continued)

V.	GENERAL DISCUSSION .....	65
	Limitations .....	68
	Conclusion .....	69
	APPENDICES .....	71
A:	Cosmetics Conditions Presented in Pilot Study Two and Study Three .....	72
B:	Interview Questions, Rating Scales, and Candidate Responses .....	74
C:	Measures .....	78
D:	Information Letters and Debriefings .....	80
E:	Video URLs .....	85
	REFERENCES .....	101

## LIST OF TABLES

Table		Page
1	Pilot Study 2 - Descriptive statistics for the dependent variable, “The candidate's makeup was applied well.” .....	87
2	Pilot Study 2 - Descriptive statistics for the dependent variable, “It was obvious that the candidate was wearing makeup.” .....	88
3	Pilot Study 2 - Descriptive statistics for the dependent variable, “How much makeup would you say the candidate was wearing?” .....	89
4	Study 3 - Descriptive statistics for all continuous variables .....	90
5	Study 3 - Correlations between continuous variables .....	91
6	Study 3 - Mean interview ratings by cosmetics condition .....	92
7	Research Question 1 - Interview rating by performance and cosmetics conditions .....	93



## LIST OF FIGURES

Figure		Page
1	Interview structure levels from Huffcutt and Arthur (1994). .....	94
2	Huffcutt et al. (2011) Model of interviewee performance as a mediating construct between candidate attributes and interview ratings.....	95
3	Hayes and Preacher (2014) mediation model with a multicategorical independent variable .....	96
4	Hypothesis 1c mediation model.....	97
5	Hypothesis 2b mediation model.....	98
6	Hypothesis 3b mediation model.....	99
7	Interaction between performance and cosmetics on interview scores.....	100

## **CHAPTER ONE**

### **INTRODUCTION**

The employment interview has been a topic of academic interest for over one hundred years, with the earliest research published in 1915 (Scott, 1915). Since then, hundreds of academic articles dedicated to further understanding the employment interview have been published. In those studies, the exact details and connotation of the employment interview has varied, so I will start by offering the definition of employment interview that will be used throughout this literature review. According to Eder and Harris (1999), the employment interview is defined as:

“...an applicant exchange of information in which the interviewer(s) inquire into the applicant’s (a) work-related knowledge, skills, and abilities (KSAs); (b) motivations; (c) values; and (d) reliability with the overall staffing goals of attracting, selecting, and retaining a highly competent and productive workforce.”  
(p. 2)

To date, the interview is the most common technique used to select candidates for employment (Eder & Harris, 1999). Interviews are the preferred selection technique among supervisors (Lievens, Highhouse, De Corte, 2005) and Human Resources (HR) practitioners (Topor, Colarelli, & Han, 2007) with good reason. First, the structured employment interview has consistently held up against legal scrutiny and is considered a legally defensible selection method (Williamson, Campion, Malos, Roehling, & Campion, 1997). Candidates tend to perceive interviews more favorably than other

selection techniques (Hausknecht, Day, & Thomas, 2004). Further, the employment interview is typically faster to develop than other types of selection assessments, which has allowed the interview to gain popularity with hiring managers (Lievens, Highhouse, & De Corte, 2005) and Human Resources professionals (Topor, Colarelli, & Han, 2007). Lastly, there is ample evidence supporting the criterion-related validity of the structured employment interview (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Schmidt & Hunter, 1998; Wiesner & Cronshaw, 1998), with fewer subgroup differences than other commonly used selection methods (Huffcutt & Roth, 1998).

Given the importance of the employment interview to obtaining employment, there is ample advice available on how to best perform during the interview, including guidance geared toward clothing and grooming. A Google search for “interview makeup” yielded over 37 million results filled with advice on how to “get hired on the spot” (Del Russo, 2015) and “nail your dream job” (Cardellino, 2013). However, there has been no empirical research to date examining the extent that facial cosmetics influence employment interview ratings. Although the empirical literature is sparse, many journalists have tackled this topic. In the article titled “The Makeup Tax,” Khazan (2015) noted that “nothing ruins a first impression like a norm violation” and “for many of us, showing up at the office or a bar without at least a swipe of blush and some mascara results in a day spent being asked if we have the flu.”

### **Study Contributions**

This study contributes to furthering the understanding of how facial cosmetics and interview performance affect employment interview ratings. Specifically, the current

study examined the effects of varying levels of facial cosmetics and interviewee performance on interview ratings. An experimental, three (no cosmetics, low cosmetics, high cosmetics) by three (low performance, intermediate performance, high performance) design was used to examine: a) what amount of facial cosmetics is most beneficial to interview ratings, b) the explanatory mediators of the cosmetics-interview ratings relationship, and c) the influence of interview performance on the cosmetics-interview ratings relationship.

The current study makes several important contributions. First, this study makes an important practical contribution for women attending employment interviews. Women have long been advised that wearing makeup to employment interviews is critical, but there is currently no empirical evidence to support this claim. This dissertation examined whether this widely accepted, yet anecdotal, advice is warranted by informing the extent to which facial cosmetics influence structured interview ratings and what amount of facial cosmetics use is most beneficial or potentially harmful to candidates during the interview. This knowledge allows women to make cosmetics choices that will be most beneficial to their future employment prospects.

In addition to practical contributions, this study also contributes to the employment interview literature. Although previous work has demonstrated the relationship between appearance and interview ratings (Barrick, Shaffer, & Degross, 2009), to my knowledge, this is the first study to examine the specific effects of facial cosmetics on structured interview ratings. Unlike attractiveness, facial cosmetics are a controllable aspect of appearance and may influence interview ratings through avenues

outside of attractiveness. This study will determine if there is a link between the use of facial cosmetics and professional appearance and perceived competence, and if differing amounts of cosmetics are perceived to be more or less professional.

Further, the experimental design of this study allows for causal relationships to be examined, which is uncommon in the industrial-organizational psychology literature. Both facial cosmetics and interviewee performance will be manipulated in order to determine their influence on employment interview ratings. The manipulation of interviewee performance is a methodological advantage to this study that allows for a better understanding of the extent that rater bias/error is present in interview ratings. Each performance level was associated with a “known” score on the interview rating scale. This was accomplished by crafting interview answers that were representative of low, intermediate, and high performance levels on the rating scale. This design simplified bias detection, as interview ratings that deviated from the known score were apparent.

Lastly, unlike other studies of facial cosmetics which have used still photographs, the current study used videos. According to media richness theory (Fletcher & Major, 2006), communication media vary in the amount of information conveyed, where face-to-face interactions provide the “richest” form of communication because of the presence of verbal (spoken word), nonverbal (hand gestures, smiling), and paraverbal (vocal inflection and tone) cues. Therefore, the use of video is advantageous as video conveys verbal and nonverbal cues that are lacking in photographs. This should lead to more accurate perceptions of appearance than photographs alone.

## **Interview Psychometric Properties**

In order to hypothesize how interview ratings are affected by cosmetics, it is first important to understand what is known about the employment interview. This literature review begins by overviewing the psychometric properties of the employment interview including interview structure, interrater reliability, criterion-related validity, incremental validity, and construct validity of the employment interview. Then, I review the factors that influence interview ratings including interviewee performance, interviewer information processing effects, and interview design considerations. Lastly, I review the literature related to appearance and interview ratings and present the study hypotheses.

**Interview Structure.** Before reviewing the research on the reliability and validity of the employment interview, it is imperative to first define structure in regards to the employment interview as structured and unstructured interviews have very different psychometric properties. Contrary to what the name suggests, interviews cannot be dichotomously defined as either structured or unstructured. Rather, interviews vary in their degree of structure depending on the specific procedures followed. Huffcutt and Arthur (1994) defined interview structure as, “the reduction in procedural variance across applicants, which can translate into the degree of discretion that an interviewer is allowed in conducting the interview” (p. 186). Campion, Palmer, and Campion (1997) offer a broader definition, defining structure as “any enhancement of the interview that is intended to increase psychometric properties by increasing standardization or otherwise assisting the interviewer in determining what questions to ask or how to evaluate responses” (p. 656).

Campion et al. (1997) proposed a comprehensive typology of interview structure that included fifteen different ways an interview can be structured along two dimensions: content and evaluation. The components of structure that influence the content of the interview include: building interview questions from job analysis, asking the same questions of all candidates, limiting follow-up questions, using better types of questions (e.g., past behavior or situational questions), using more interview questions, controlling ancillary information, and not allowing questions from the candidate until after the interview. The components of structure that influence the evaluation of the interview include: rating each answer, using detailed anchored rating scales, taking detailed notes, using multiple interviewers, using the same interviewers across all candidates, not discussing candidates between interviews, training interviewers, and using statistical prediction methods.

In an effort to examine the underlying factor structure of interview structure components, Chapman and Zweig (2005) surveyed 1,500 employers who conducted interviews with approximately 4,000 applicants on the different elements of interview structure identified by Campion et al. (1997). These data were analyzed using principal axis factor analysis, which is an exploratory technique, and a three-factor solution emerged. These factors accounted for 43.45% of the variance in items and were named evaluation standardization, question sophistication, and question consistency.

More recently, researchers have suggested additional components that could increase interview structure. Levashina, Hartwell, Morgeson, and Campion (2014) suggest that limited rapport building, interview transparency, and recording of interviews

may also be elements of interview structure. Limiting the rapport building or “small talk” that occurs prior to the interview is thought to increase interview structure by preventing interviewers from forming early impressions on information irrelevant to the job.

Interview transparency is the degree to which interviewees are informed of the constructs being assessed by the interview. The authors suggest that when interviews are transparent, candidates can share more relevant past experiences in their answers. Lastly, recording of interviews is hypothesized to provide additional accountability for interviewers, resulting in more accurate ratings and therefore improved interview reliability and validity.

Another recent study (Hartwell & Campion, 2016) found that giving interviewers normative feedback, which consisted of showing individual interviewers how their interview ratings compared to others’ ratings, resulted in broader usage of the rating scale and encouraged lenient and severe raters to adjust their ratings. Additionally, after interviewers received normative feedback, interrater reliability and interrater agreement improved. This suggests that giving interviewers normative feedback on their ratings may also be considered a component of interview structure.

**Interrater Reliability.** Three types of measurement error that threaten the interrater reliability of employment interviews have been identified in the literature: random response error, transient error, and conspect error (Huffcutt, Culbertson, & Weyhrauch, 2013; Schmidt & Zimmerman, 2004). Measurement error is broadly defined as the difference between an observed score and an actual or true score. Random response error is the variation in candidate responses to the same basic question within



the same interview and is caused by fluctuations in the candidate's attention and mental resources. Transient error is variation in candidate responses to the same question across different interviews, while conspect error is variation in interviewer ratings of the same candidate response. The only interview format that accounts for all three types of identified measurement error is separate interviews conducted by different interviewers.

Huffcutt et al. (2013) meta-analyzed 125 interrater reliability coefficients. Studies were coded by level of structure and format (panel or separate interviewers). Results indicated that interrater reliability increased as structure increased ( $M = .36$  at level one,  $M = .51$  at level two,  $M = .65$  at level three,  $M = .79$  at level four, and  $M = .76$  at level five, where level one has the least structure and level five has the most structure). Further, panel interviews ( $M = .78$ ) had higher interrater reliability than separate interviews ( $M = .44$ ). The authors suggest that this is because panel interviews do not account for the types of measurement error listed above, leading to overestimation of interrater reliability.

**Criterion-Related Validity.** Early meta-analytic evidence suggested the employment interview had relatively low criterion-related validity (.14) compared to other types of assessments (Hunter & Hunter, 1984); however, this early evidence did not distinguish between structured and unstructured interviews. Today, there is ample evidence that adding structure to employment interviews greatly improves their psychometric properties. Several meta-analyses have demonstrated that structured interviews have superior reliability and validity to unstructured interviews (Conway, Jako, Goodman, 1995; Huffcutt & Arthur, 1994; Huffcutt et al., 2013; Huffcutt,

Culbertson, & Weyhrauch, 2014; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Schmidt & Hunter, 1998; Wiesner & Cronshaw, 1998). Criterion-related validity estimates for unstructured interviews range from .20 to .34 (Wiesner & Cronshaw, 1988; McDaniel et al., 1994), while corrected estimates for structured interviews have been as high as .71 (Huffcutt et al., 2014).

Huffcutt and Arthur (1994) suggested that the level of interview structure moderates criterion-related validity. The authors identified four levels of interview structure based on the standardization of interview questions and response scoring. The first level imposed no constraints on questions and used a global scoring method, which also describes an unstructured interview. The second level posed some constraints on scoring or question standardization, while level three involved complete question standardization. Lastly, the fourth level was the highest level of structure and consisted of predetermined questions, ratings of each question, and the use of benchmarks to assist scoring. See Figure 1 for more details on each level. Huffcutt and Arthur (1994) meta-analyzed 114 validity coefficients and, after correcting for range restriction and unreliability in the criterion, found  $\rho = .20$  for level one,  $\rho = .35$  for level two,  $\rho = .56$  for level three, and  $\rho = .57$  for level four. More recently, Huffcutt et al. (2014) provided an update to the Huffcutt and Arthur (1994) meta-analytic estimates using new techniques to correct for indirect range restriction. Estimates of criterion-related validity were found to be considerably higher than previously thought:  $\rho = .20$  at level one,  $\rho = .46$  at level two,  $\rho = .71$  at level three, and at  $\rho = .70$  at level four. These results demonstrate that structure moderates the relationship between interview scores and performance ratings such that

increased structure results in higher validity coefficients; however, there may be a point at which increased structure does not further enhance validity (e.g., level three vs. level four).

Researchers have offered multiple explanations for why increased structure improves the validity of interviews. A common explanation is that structured interviews are more reliable than unstructured interviews (Conway et al. 1995; Huffcutt et al. 2013; Schmidt & Zimmerman, 2004). It has also been suggested that adding structure to the method of interview evaluation reduces the cognitive complexity of response processing (Tsai, Chen, & Chiu, 2005), focuses the attention of the interviewer on job-related content (Huffcutt et al., 2011), and reduces bias (Lin, Dobbins, & Farh, 1992).

**Incremental Validity.** Incremental validity of the structured interview is the extent that the interview predicts the criterion - typically job performance - over and above other predictors. Because cognitive ability is generally considered the best predictor of job performance (Hunter & Hunter, 1984), the interview has been most commonly assessed relative to measures of cognitive ability. Previous research on the incremental validity of employment interviews has yielded mixed results. Campion, Pursell, and Brown (1988) found that structured interviews did not explain additional variance beyond a battery of cognitive ability tests. Similarly, studies by Shahani, Dipboye, and Gehrlein (1991) and Walters, Miller, and Ree (1993) both found evidence for the validity of the interview, but failed to find evidence of incremental validity above other predictors.

Conversely, several studies found that structured interviews do provide incremental validity over and above cognitive ability (Campion, Campion, & Hudson, 1994; Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Pulakos & Schmitt, 1995; Schmidt & Hunter, 1998). Campion et al. (1994) examined the incremental validity of structured interviews by question type (experience based or situational) beyond a battery of cognitive ability tests. Results indicated that both question types predicted job performance over and above the test battery.

Cortina et al. (2000) found evidence that interviews predict job performance over and above cognitive ability and conscientiousness. Unstructured (level one) interviews accounted for an additional 1.5% of the variance, structured interviews at level two accounted for an additional 3.7% of the variance, and structured interviews at levels three and four accounted for an additional 16.9% of the variance above cognitive ability and conscientiousness in predicting job performance. These results indicate that additional interview structure results in higher incremental validity over and above other predictors. The authors suggest that previous work that did not find support for incremental validity of the interview may be due to lower levels of interview structure found in those studies.

**Construct Validity.** Although there is compelling evidence that structured employment interviews predict job performance, it is less clear what constructs are assessed by the interview and why structured interviews predict performance. It is important to emphasize that the employment interview can be designed to measure various constructs, as the employment interview is a method, not a construct (Arthur & Villado, 2008). Huffcutt, Conway, Roth, and Stone (2001) examined the constructs that

were assessed in 338 interview dimensions across 47 studies. Through a review of the extant literature, the authors created a taxonomy of seven major categories by which interview constructs were grouped: mental capability, knowledge and skills, personality tendencies, applied social skills, interests and preferences, organizational fit, and physical attributes. The most commonly measured construct category was personality tendencies (35%), followed by applied social skills (28%), mental capability (16%), knowledge and skills (10%), interest and preferences (4%), physical attributes (4%), and lastly, organizational fit (3%).

The work of Huffcutt et al. (2001) provides important insight into the types of constructs that interviews are designed to measure, but does not address how well interviews actually assess these constructs. Much of the research on the construct validity of employment interviews has been inconclusive. Multiple studies have failed to find support for the construct validity of interviews (Conway & Peneno, 1999; Huffcutt, Weekley, Wiesner, Degroot, & Jones, 2001; Van Iddekinge, Raymark, Eidson, & Attenweiler, 2004), while others have demonstrated that interviews can be effective measures of different constructs (Klehe, Konig, Richter, Kleinmann, & Melchers, 2008; Van Iddekinge, Raymark, & Roth, 2005).

Van Iddekinge et al. (2004) expanded on the work of Huffcutt et al. (2001) by examining the construct validity of two behavioral description interviews (BDIs) used to select customer service managers for a large grocery organization. Between the two interviews (A and B), there were six items that measured interpersonal skills, four items that measured conscientiousness, and four items that measured stress management.

Confirmatory factor analysis indicated that the constructs the interviews were meant to measure accounted for only 9.3% of the variance. Likewise, the convergent validity of the interview with tests designed to measure the same general construct was .05 for interview A and .04 for interview B. The discriminant validities were also very small, .04 and .03 respectively. The authors concluded that:

“..the impressive criterion-related validity of BDIs may not be due to good construct-related validity. Instead, structured behavioral interviews may derive their predictive validity from certain interviewee characteristics (unrelated to the constructs of interest) that can be reliably assessed in an interview setting” (p. 86-87).

A similar study by Van Iddekinge et al. (2005) found more compelling evidence of the construct validity of an interview designed to measure personality traits. In mock interviews, interviewees were asked three questions about each personality construct, for a total of nine items. Interviewees also completed validated, self-report assessments of the same personality traits. Multitrait-multimethod analyses revealed good convergent and discriminant validity, and a confirmatory factor analysis of the interview ratings showed acceptable fit for a three-factor model of the interview ratings. These results provide support for the construct validity of the interview used in this study.

Researchers have also examined possible moderators of interview construct validity. Klehe et al. (2008) examined the impact of interview transparency on both construct and criterion-related validity of the interview. Transparency is the extent that interviewees are informed of the criteria being evaluated in the interview. The authors

point out that candidates who misjudge the intent or purpose of a question are less likely to give examples that represent the construct being assessed, and will subsequently perform poorly compared to candidates who fully understood the item, regardless of the candidate's true level of the intended construct. Subsequently, items carefully designed to measure the same construct may still fail to converge in nontransparent interviews. Klehe et al. (2008) assigned participants and interviewers to either a transparent or nontransparent interview condition. Those in the transparent condition received a handout with the general dimensions that would be examined in the interview. Participants in the transparent interview condition received higher interview scores than those in the nontransparent condition on all constructs measured. Further, the transparent condition also demonstrated a slight increase in construct validity. Interestingly, transparency did not have an effect on the criterion-related validity of the interview.

Similar to interview transparency, another factor that has recently been examined is the candidate's ability to identify the criteria (ATIC) being measured by the interview. Melchers et al. (2009) suggest that ATIC will be positively related to interview construct validity because candidates who can correctly identify the dimension(s) measured will be able to describe more relevant experiences and behaviors to that construct. For example, if a candidate incorrectly identifies the criteria being measured as assertiveness when the actual criteria being measured is cooperation, that candidate will describe behaviors unrelated to cooperation and score low on that question.

In order to determine the impact of ATIC on interview construct validity, Melcher et al. (2009) interviewed 92 participants using an interview designed to measure three

different constructs. After the interview was complete, candidates were given a questionnaire with all of the interview questions. They were then asked to write down what they had thought each question was trying to assess during the interview. These hypotheses were then coded for accuracy. A multitrait-multimethod analysis indicated that the entire interview had poor construct-related validity. However, when only items that were correctly identified were included in the analysis, construct-related validity improved. This indicates that when candidates correctly identify the criteria being measured in an interview question, ratings are more valid.

**Summary of Interview Psychometric Properties.** To summarize, the psychometric properties of the employment interview have been well-researched. The most consistent and robust finding is that structured interviews have better psychometric properties than unstructured interviews. For example, interrater reliability (Huffcutt et al., 2013), criterion-related validity (Huffcutt et al., 2014), and incremental validity (Cortina et al., 2000) increase with additional structure. The research regarding the construct validity of the interview has been less clear as some studies have found support for the construct validity of the interview (Klehe et al., 2008; Van Iddekinge et al., 2005), while others have not (Conway & Peneno, 1999; Van Iddekinge et al., 2004). Overall, the psychometric properties of the employment interview indicate that it is a useful selection practice. Therefore, it is important to further understand the factors that influence interview ratings.

### **Interview Ratings**



The interview is presumed to function as an assessment tool to gauge an interviewee's level of a specific construct or characteristic in order to distinguish between candidates who have high or low levels of that construct or characteristic. This is done by having one or more interviewers rate candidates' answers to a series of questions. However, general understanding of the many factors that influence how interview ratings are made and the accuracy of those ratings is limited. Huffcutt, Van Iddekinge, and Roth (2011) explain that this is likely because research has instead concentrated on the properties of interview ratings, such as reliability (Conway et al., 1995) or criterion-related validity (Huffcutt et al., 2014). Recently, Huffcutt et al. (2011) proposed a model of the constructs that influence the ratings an interviewer gives an interviewee (see Figure 2 for the complete model). The authors hypothesized that the proximal predictors of interview ratings are interviewee performance, interviewer information processing effects, and interview design considerations.

**Interviewee performance.** Huffcutt et al. (2011) define interviewee performance as the capability of interviewees to present their qualifications during the interview, or what the interviewee says and does during the interview. More specifically, interviewee performance consists of three subcategories: interviewee answers to the questions, delivery of interview answers (e.g., articulation and pitch), and the nonverbal behaviors of the interviewee, such as facial expressions, posture, and appearance. The model proposes that interviewee performance is predicted by three proximal sets of factors: core candidate qualifications, interviewee state influences, and interviewer-interviewee dynamics. Core candidate qualifications are defined as the interviewee's declarative

knowledge, procedural knowledge and skills, and motivation, and are a result of interviewees' more general attributes. The general attributes include mental ability; personality; education, training, and experience; and interests, goals, and values. Interviewee state influences include interview self-efficacy, interviewing motivation, and interview anxiety. Lastly, Interviewer-Interviewee Dynamics consists of interviewee social effectiveness and interviewer personality.

Ideally, candidates would always answer interview questions to the best of their ability. However, Huffcutt et al. (2011) note that candidates do not always accurately demonstrate their true ability level during an interview due to individual differences and the complex social requirements of the situation. The authors posit that interviewee performance is a primary mediating factor between candidate characteristics (e.g., declarative knowledge, procedural skills and abilities, motivation, mental ability, personality) and interviewer ratings.

Candidates frequently use impression management techniques to bolster their perceived performance during the interview. Ellis, West, Ryan, and DeShon (2002) found that nearly all candidates used some form of impression management during structured interviews. The types of impression management tactics (IM) used have been divided into verbal and nonverbal IM. Nonverbal IM includes behaviors such as making eye contact and smiling. Verbal IM has been divided into two categories: assertive and defensive. Assertive IM are techniques used to proactively create a positive image, such as self-promotion, ingratiation, or opinion conformity. Lastly, defensive impression management techniques can be used to repair negative perceptions (e.g., apologizing).

Levashina et al. (2014) meta-analyzed the effect of the different impression management techniques on interview ratings and found that self-promotion had the strongest effect ( $r = .26$ ), followed by nonverbal IM ( $r = .18$ ), other-focused IM ( $r = .13$ ), and defensive IM ( $r = .12$ ). These results provide additional evidence that the performance of the interviewee influence interview ratings.

Although it may seem intuitive that interviewee performance would impact interview ratings, there have not been any experimental studies that held nonverbal behaviors and delivery of the answer constant to independently determine the effects of what the candidates says on interview ratings. Therefore, based on the previous research and the theoretical work of Huffcutt et al. (2011), a pilot study will be done to ensure there is a main effect of interviewee performance on interview ratings.

**Interviewer information processing effects.** In addition to interviewee performance, another primary factor influencing interview ratings is the ability of the interviewer to make valid ratings/judgments about the interviewee's answers. The shortcomings of the interviewer have been demonstrated in the research literature by the ample evidence that interview ratings are susceptible to bias. Bias is defined as "systematic group differences in item responses, test scores, or other assessments for reasons unrelated to the trait being assessed" (Highhouse, Doverspike, & Guion, 2016, p. 172). Experiments utilizing the same confederate in pregnant and nonpregnant conditions have shown that pregnant interviewees receive lower interview ratings and are less likely to be recommended for hire (Bragger, Kutcher, Morgan, & Firth., 2002; Cunningham & Macon, 2007). This bias may be due to perceived incompetence, lack of commitment,

inflexibility, or the need for accommodation (Morgan, Walker, Hebl, & King, 2013). Relatedly, being overweight (Pingitore, Dugoni, Tindale, & Spring, 1994) and speaking with an accent (Deprez & Sims, 2013; Purkiss, Perrewe, Gillespie, Mayes, & Ferris, 2006) have also been related to lower interview ratings. However, group differences have not been found for demographic characteristics including race and gender (Levashina et al., 2014). In addition, interviews are also susceptible to rating errors including halo (Crissy & Regan, 1951), leniency (O'Brien & Rothstein, 2011), and contrast effects (Wexley, Sanders, & Yukl, 1973).

The tendency for humans to make systematic errors in judgment has long been recognized by psychologists. The large cognitive demand placed on interviewers may encourage the use of simplifying heuristics and increase the likelihood of rating errors during the interview (Huffcutt et al., 2011), resulting in less valid and reliable ratings. For example, due to memory limitations, the interviewer may only use the information they are able to recall to make ratings, which means a potentially large amount of information about the candidate is not considered when ratings are made. Further, research has shown that interview ratings are highly influenced by interviewer initial impressions, even in highly structured interviews (Barrick et al., 2012; Barrick, Swider, & Stewart, 2010). However, it is important to note that interview decisions are not typically “snap judgments” made by the interviewer (Frieder, Van Iddekinge, & Raymark, 2015).

Interview bias has also been explained in the context of dual-process theory (Derous, Buijsrogge, Roulin, & Duyk, 2016). This theory proposes that human decision-making is driven by two processes referred to as System 1 and System 2 (Kahneman &

Frederick, 2002; Stanovich, 1999). System 1 is driven by quick, automatic reactions that rely heavily on biases and heuristics, while System 2 processes are slow and deliberative (Evans, 2008). Derous et al. (2016) suggest that interviewers' initial impressions are based on System 1 processing, and therefore are frequently biased and incorrect. These quick, initial reactions are coupled with the high cognitive demands of the interviewing situation and result in interviewers who are unable to correct their initial impressions. Final decisions are ultimately anchored in the initial bias of the interviewer, thus resulting in biased and inaccurate ratings. Moreover, the problems surrounding these biased ratings are further complicated by individuals' inability to detect bias in their own ratings (Pronin, 2007; Pronin & Ross, 2002), an effect referred to as the bias blind spot.

**Interview design considerations.** The last proximal predictor proposed by Huffcutt et al. (2011) of interview ratings is a set of interview design considerations, including level of interview structure and the amount of pre-interview information available to the interviewer. Huffcutt et al. suggested that both additional interview structure and reducing the amount of pre-interview information available to the interviewer should improve the quality of interview ratings by reducing the saliency of any extant interviewer information processing effects. For example, if the interviewer takes detailed notes on what the candidate is saying during the interview and rates each question (versus giving a global score for the entire interview), memory limitations should have less impact on the ratings given by that interviewer. Additionally, if behaviorally anchored rating scales are used to rate each interview question, the interviewer should have ample job-related guidance to use when rating answers, thereby

reducing his or her reliance on simplifying heuristics and susceptibility to various biases and errors.

### **Appearance**

Overall, the work of Huffcutt et al. (2011) provides a useful framework of the variables and constructs that influence interviewer ratings. In addition to the three proximal predictors of interviewer ratings discussed above, another important factor to consider is interviewee appearance. The model proposed by Huffcutt et al. (2011) hypothesizes that personal/demographic characteristics, including the attractiveness (i.e. appearance) of the applicant, is a distal predictor of interview ratings as there is evidence that appearance is an important predictor of interview ratings.

Motwidlo and Burnett (1995) investigated how interview ratings made based only on visual cues (recorded video of candidates with no sound) correlated with performance. They found that when raters could see and hear the candidate, ratings correlated .36 with performance. Interestingly, when ratings were made based only on sight, ratings correlated .32 with performance. This indicates that visual cues (e.g., appearance, clothing, hand gestures, smiling) may account for actual variance in performance. Alternatively, it is also possible that supervisor ratings of performance are subject to the same systematic error as interview ratings based solely on sight.

Posthuma, Morgeson, and Campion (2002) called researchers to examine the differential effects of controllable versus uncontrollable facets of physical appearance, suggesting that when applicants do something to improve their appearance (e.g., appropriate clothing and grooming) interviewers may react more favorably. An

experimental study by Mack and Rainey (1990) found that participants were more likely to report that they would hire well-groomed candidates than poorly-groomed candidates with the same job qualifications. Similarly, Kinicki and Lockwood (1985) found that recruiters relied primarily on subjective criteria such as appearance and attraction when making employment recommendations.

A recent finding from the sociological literature lends further support to the idea that enhancing controllable aspects of one's appearance is advantageous in the workplace. Wong and Penner (2016) note that attractiveness is typically thought of as a biologically fixed trait, however attractiveness may be more appropriately described as a combination of biological traits, personality characteristics, and beauty practices. These beauty practices, or "beauty work" as they are described in the sociological literature, consist of practices such as wearing cosmetics, exercising/dieting, wearing stylish clothing, or even getting plastic surgery, all with the goal of improving appearance. Supporting the results of previous studies, Wong and Penner (2016) found that there was a significant attractiveness-based difference in income, such that more attractive people earn about 20% more than people of average attractiveness. Interestingly, when the authors added grooming to the regression model, the difference between more attractive women and women of average attractiveness became statistically insignificant, which suggests that grooming may be more important than attractiveness in predicting salary. The authors suggest that grooming practices and putting effort into one's appearance may signal willingness to put effort into other activities, such as work.

**Influence of cosmetics on appearance.** One controllable aspects of appearance that has not been empirically examined in the employment interview literature is the use of facial cosmetics. Cosmetics have been used for thousands of years across a multitude of cultures (Eldridge, 2015). One society most noted for their use of cosmetics is the ancient Egyptians who used kohl to create heavy lines around the eyes (Eldridge, 2015). Even when cosmetics were banned during the Victorian era, women still found ways to alter their appearance through pinching their cheeks and biting their lips to increase coloration (Peiss, 1998). When the concentration camp at Bergen-Belsen was liberated in April of 1945, a large shipment of lipstick was sent to the internees. About this event, Lieutenant Colonel Mervin Willett Gonin wrote,

“I wish so much that I could discover who did it, it was the action of genius, sheer unadulterated brilliance. I believe nothing did more for those internees than the lipstick. ...At last someone had done something to make them individuals again, they were someone, no longer merely the number tattooed on the arm. At last they could take an interest in their appearance. That lipstick started to give them back their humanity.” (Gonin, 1945).

Today, the use of cosmetics is still widespread among women and is growing amongst men (Squier, 2016; Whipp, 2017). Russell (2010) posits that this is because cosmetics “...are a part of what defines us individuals and as humans.” (p. 186).

Cosmetics have also been shown to improve the quality of life for pediatric patients with visible skin anomalies (Ramien, et al., 2014). Children between the ages of 5 and 18 years old that had visible skin conditions were taught how to camouflage their



conditions with cosmetics by a trained makeup artist. Results of one and six month follow up surveys indicated that quality of life improved for these patients. The authors note that cosmetics may be particularly beneficial for patients in helping patients with new diagnoses adapt and avoid anxiety related to their skin problem.

*Effect of cosmetics on physical attractiveness.* Research has consistently demonstrated a positive relationship between the use of cosmetics and ratings of physical attractiveness (Cash, Dawson, Davis, Bowen, & Galumbeck, 1989; Cox & Glick, 1986; Etcoff, Stock, Haley, Vickery, & House, 2011; Miller & Cox, 1982; Mulhern, Fieldman, Hussey, Leveque, & Pineau, 2003; Workman & Johnson, 1991). The majority of studies examining the influence of cosmetics on attractiveness used photographs of women either wearing or not wearing cosmetics and asked participants to rate attractiveness. To test the effect of cosmetics in a natural setting, Gueguen (2008) measured the number of men that approached female confederates in a bar. When the female confederates were wearing makeup, they were approached by significantly more men than when they were not wearing makeup. Further, the amount of time before verbal contact was made with the confederate was significantly shorter in the cosmetics condition.

In an effort to determine if different types of cosmetics have differential effects on attractiveness, Mulhern et al. (2003) asked participants to view ten sets of five photographs and rank each set from most attractive to least attractive. Each set of photographs consisted of the same volunteer with varying levels of makeup: no makeup, foundation only, eye makeup only, lip makeup only, and full facial makeup. Both men and women rated the “full face” makeup photographs as the most attractive, followed by

the eye makeup only and foundation only conditions. Contrary to expectations, lipstick did not independently increase attractiveness ratings.

There are many reasons that account for the increase in attractiveness perceptions when cosmetics are used. First, cosmetics have been shown to increase the contrast between facial features and the skin, which increases the perceived femininity of the face and increases attractiveness (Jones, Russell, & Ward, 2015; Russell, 2010). Cosmetics may also enhance facial symmetry (Gold, 2011), which is related to increased attractiveness (Perrett et. al, 1999) and hide facial imperfections (Fink, Grammer, & Thornhill, 2001; Ramien et al., 2014). Given the previous research that indicates cosmetics have a positive influence on ratings of physical attractiveness, the following hypothesis is proposed:

*Hypothesis 1a:* Ratings of interviewee physical attractiveness will be higher in the cosmetics conditions than in the no-cosmetics condition.

The advantages associated with attractiveness have been well documented in the social psychology literature. Given the abundance of research on this topic, multiple meta-analyses have synthesized the research findings (Barrick et al., 2009; Eagly, Ashmore, Makhijani, & Longo, 1999; Feingold, 1992; Hosoda, Stone-Romero, & Coasts, 2003; Jackson, Hunter, & Hodge, 1995; Langlois et al., 2000). Jackson et al. (1995) examined the influence of attractiveness on intellectual competence. Results indicated that attractive individuals are perceived to be more intellectually competent than unattractive individuals. The primary explanation for the influence of attractiveness on various outcomes is the “what is beautiful is good” stereotype which is the tendency for

positive traits to be attributed to attractive people, while negative traits are attributed to unattractive people (Dion, Berscheid, & Walster, 1974).

Langlois et al. (2000) found that attractive individuals were judged to have greater occupational competence, interpersonal competence, and social appeal. These judgments were defined as “informed opinions about attractive and unattractive targets...based on actual incidents of observable behavior” (p. 397). Moreover, results indicated that attractive individuals are treated significantly better than unattractive individuals. For example, attractive individuals are given more attention, help/cooperation, and rewards. Lastly, the authors examined behavior differences between attractive and unattractive individuals. Attractive individuals experienced more occupational success, had more dating and sexual experience, better physical health, were more extraverted, and had better mental health than unattractive individuals. Ultimately, this research indicates that attractiveness has a meaningful influence on daily life and is not limited to mere impressions.

Extending the work of Langlois et al. (2000), Hosoda et al. (2003) meta-analyzed experimental studies that examined the influence of physical attractiveness specifically on job-related outcomes. Results followed suit with those of previous meta-analyses. The mean weighted effect size of attractiveness was .34. The outcomes examined included suitability ranking, hiring decision, promotion decision, predicted success, suitability ratings, employment potential, choice as business partner, and performance evaluation. This meta-analysis also found no differences based on the sex of the applicant/employee or the sex-type of the job, indicating that attractiveness was always a beneficial trait.

Further, the influence of attractiveness was not significantly different in the presence or absence of job-related information.

A meta-analysis by Barrick et al. (2009) examined the relationship between self-presentation tactics (i.e. appearance, impression management, verbal and nonverbal behavior) and interview ratings. The authors categorized appearance into two types: physical and professional. Physical appearance was described as the beauty or physical appeal of an individual, while professional appearance was described as an individual's hygiene, grooming, and clothing. Of the three self-presentation tactics examined, appearance had the strongest relationship with interview ratings ( $r = .53$ ). The corrected sample-weighted mean correlation was .54 between physical appearance and interview ratings and was .48 between professional appearance and interview ratings. Analyses also indicated that as interview structure increased, the relationship between appearance and interview scores decreased. However, even when interviews were highly structured (level 3), there was still a small, but significant relationship between appearance and interview scores ( $r = .18$ ).

Because cosmetics are hypothesized to increase attractiveness, and attractiveness has been associated with higher ratings of competence and higher interview ratings, the following hypotheses are proposed:

*Hypothesis 1b:* Interview ratings will be higher in the cosmetics conditions than in the no cosmetics condition.

*Hypothesis 1c:* Ratings of interviewee physical attractiveness will mediate the relationship between cosmetics and interview ratings.

*Effect of cosmetics on perceived competence and professional appearance.*

Today, cosmetics are primarily used by women to promote a positive image (Robertson, Fieldman, & Hussey, 2008). Although cosmetics were not specifically examined, Ruetzler, Taylor, Reynolds, Baker, and Killen (2012) found that the largest contributing factor to professional appearance is being neatly groomed, which cosmetics could influence. Dellinger and Williams (1997) performed in-depth interviews with 20 women about their use of makeup at work. Of the 20 women interviewed, 14 reported that they wear makeup every day at work. Women reported that they wear makeup to feel more polished and confident at work. One of the major concerns about not wearing makeup to work was that they would be perceived as less credible without makeup. One woman said about makeup, “It’s one of the things you do to excel...I’ve seen female attorneys go to court and looked washed out and people just do not react as positively...” (p. 165). This indicates that wearing makeup at work may be an important part of gaining respect and looking professional.

Although there is no research that directly examines the link between cosmetics and professional appearance, there is evidence that the use of cosmetics may relate to increased perceptions of constructs related to professionalism. For example, Etkoff, Stock, Haley, Vickery, and House (2011) found that women wearing cosmetics were perceived as more competent. Nash, Fieldman, Hussey, Lévêque, and Pineau (2006) examined the relationship between use of cosmetics and participants’ perceptions of four constructs: health, confidence, future earning potential, and professional status. The authors presented participants with photos of women with or without makeup. Results

indicated that participants rated the women wearing cosmetics as healthier, more confident, and as having greater future earning potential than women who were not wearing makeup. Additionally, women wearing cosmetics were more likely to be categorized into a high or average professional status, and women who were not wearing makeup were more likely to be categorized into a low-status profession or as unemployed. Another study found that the use of cosmetics increased perceptions of women's prestige and dominance (Mileva, Jones, Russell, & Little, 2016), which are two traits that could also be related to perceived competence and professional appearance.

Collectively, these results support the notion that the use of cosmetics is not just associated with increased attractiveness, but also with increased perceptions of professionalism and competence. Cosmetics contribute to a well-groomed appearance that signals professionalism and competence, which in turn could inflate interview ratings. The following hypotheses are proposed:

*Hypothesis 2a:* Ratings of interviewee professional appearance will be higher in the cosmetics conditions than in the no cosmetics condition.

*Hypothesis 3a:* Ratings of perceived interviewee competence will be higher in the cosmetics conditions than in the no cosmetics condition.

*Hypothesis 2b:* Ratings of interviewee professional appearance will mediate the relationship between cosmetics and interview ratings.

*Hypothesis 3b:* Perceived interviewee competence will mediate the relationship between cosmetics and interview ratings.

*Amount of cosmetics on professional appearance.* There is also evidence that the amount of cosmetics worn differentially influences perceptions. One study found that both men and women overestimate the amount of cosmetics preferred by others, suggesting that “less is more” when applying cosmetics (Jones, Kramer, & Ward, 2014). Tagai, Ohtaka, and Nittono (2016) examined ratings of attractiveness of female faces wearing no makeup, light/natural makeup, and heavy/glamorous makeup. Ratings of attractiveness were highest in the light makeup condition and lowest in the no makeup condition.

Another study presented images of two women ranging in age from 25-50 in one of four increasingly heavy makeup conditions: no makeup, natural makeup, professional makeup, and glamorous makeup (Etcoff et al., 2011). Participants were asked to rate each image on attractiveness, likeability, trustworthiness, and competence. Two analyses were performed. The first compared the aggregated makeup conditions to the no-makeup condition. Results indicated that there was a significant, positive main effect on judgments of all outcomes in the aggregated makeup condition, indicating that participants perceived women wearing makeup to be more attractive, likeable, trustworthy, and competent than women not wearing makeup. The second analysis compared the individual makeup conditions to one another. Each makeup condition also had a significant positive effect on judgments of competence and attractiveness. The natural and professional conditions both had a significant positive effect on likeability, while the glamorous condition did not have a significant effect. Interestingly, the effect of each makeup condition was different on judgments of trustworthiness. The natural

condition had a significant positive effect, the professional condition did not have a significant effect, and the glamorous condition had a significant negative effect.

Some research has found that the use of heavy makeup is associated with negative perceptions. Huguet, Croizet, and Richetin (2004), examined the effects of facial makeup on impression formation. Results indicated that makeup had positive effects on ratings of attractiveness, which is consistent with previous research; however, faces wearing heavy makeup were rated as more vain, unfaithful, and shallow than the faces without makeup. The authors note that “a woman wearing distinctive (heavier) makeup can be seen as choosing to send strong seduction signals (sexually confident/assertive and possibly unfaithful). She can also be seen as highly invested in her appearance (vain, shallow, and not so bright)...” (p. 1765).

The results of previous research indicate that the amount of makeup worn differentially affects perceptions of likeability and trustworthiness (Etoff et al., 2011). This provides some support that wearing no cosmetics and alternatively wearing “too much” makeup could be considered inappropriate for the workplace. Women who do not wear any facial cosmetics could be perceived to care little about their appearance, which could then lead to the perception that they are less conscientious or hardworking. On the other hand, wearing too much makeup could be considered distracting and unprofessional. Therefore, the following hypothesis is proposed:

*Hypothesis 2c:* Ratings of professional appearance will be highest in the low cosmetics condition, followed by the high cosmetics condition, and lowest in the no-cosmetics condition.



## **Interaction between Performance and Cosmetics**

There is evidence that the use of facial cosmetics positively influences perceptions of attractiveness (Mileva et al., 2016), competence (Etcoff et al., 2011), and professional status (Nash et al., 2006). However, these previous studies typically did not provide participants with information regarding the actual performance or ability level of the person being rated. Participants were rating their perceptions of photos; therefore, it is unclear if the effects of facial cosmetics will positively influence interview ratings at low, intermediate, and high levels of performance, or if facial cosmetics will have more influence at different levels of performance. Jackson et al. (1995) found that the effect of attractiveness was stronger when there was no accompanying information about the individual's competence than when there was competence information available. However, the biasing effect of attractiveness was still present even when competence information was present. Koch, D'Mello, and Sackett (2015) meta-analyzed the influence of gender stereotypes and bias on employment decision making and found that bias did not consistently decrease when participants were provided with additional information about candidates. However, results did indicate that when competence of the candidate was high, gender bias was reduced.

It is possible that at low levels of performance, cosmetics will have a greater influence on ratings because the candidate is presumably providing less job-relevant information to the interviewer. Some argue that stereotypes are most likely to be used when there is little relevant information available (Landy, 2008). Therefore, in the low

performance condition the interviewer may be more likely to rely on first impressions and heuristics, and make less valid judgments about the candidate as a result.

Alternatively, cosmetics could influence interview ratings very little at high and low levels of performance due to the saliency of performance level. For example, it is possible that fewer cognitive resources are required to identify the candidate's performance level when the candidate is performing very well or very poorly, thus making it easier for the interviewer to identify high and low levels of performance. If this is the case, the influence of potential appearance-based biases would be decreased when performance is either high or low.

Because there are alternative explanations for how interviewee performance will moderate the relationship between cosmetics and interview ratings, no directional hypotheses are proposed. Instead, this relationship will be explored through the following research question:

*Research Question 1: Will the influence of facial cosmetics on interview ratings vary at different levels of interview performance?*

### **The Current Study**

This dissertation was conducted over three studies. First, Pilot Study One provided a manipulation check of the interview performance variable. I propose that interviewee performance will affect interview ratings such that interview ratings will be highest in the high performance condition, followed by the acceptable performance condition, and lowest in the low performance condition. This manipulation check was done by asking participants to listen to and rate audio recording of interview questions at

either low, intermediate, or high performance. It is important to confirm that participants are able to identify differences in interview performance in order to examine how interview performance affects other variables. For example, if participants cannot reliably distinguish between low and high performance, then I would not expect to see interview performance act as a moderator of other relationships.

Next, Pilot Study Two served as a manipulation check of the cosmetics variable and to inform which cosmetics conditions to use in study three. In Pilot Study Two, participants watched a short introductory video of a candidate that is wearing one of five facial cosmetics applications ranging from no cosmetics to a very heavy, dark application of facial cosmetics. This was done to ensure that the cosmetics conditions are meaningfully different from one another. This was necessary because perceptions of facial cosmetics are subjective, and the differences between various combinations of facial cosmetics applications (i.e. makeup looks) can be subtle and difficult to identify. Likewise, Pilot Study Two also examined perceptions of makeup application quality to ensure that there were not significant differences in the perception of makeup application quality between conditions. The results of Pilot Study Two were used to determine which cosmetics stimuli to use in study three.

Lastly, study three tested Hypotheses 1-3 and Research Question 1 using a 3 (cosmetics) by 3 (interview performance) between-subjects design. The purpose of study three was to examine the effects of facial cosmetics on structured interview scores. Study three also examined how facial cosmetics relate to attractiveness, professional appearance, and perceived competence and if any of these variables mediate the

relationship between facial cosmetics and structured interview scores. Lastly, study three examined interview performance as a potential moderator of the cosmetics-interview performance relationship.

## **CHAPTER TWO**

### **PILOT STUDY ONE**

The purpose of Pilot Study One was to ensure that interviewee performance affected interview ratings such that interview ratings were highest in the high performance condition, followed by the acceptable performance condition, and lowest in the low performance condition. This pilot study served as a manipulation check of the interview performance variable by ensuring that participants were able to adequately distinguish between low, intermediate, and high performance.

#### **Method**

##### **Participants**

Participants were recruited using Amazon's Mechanical Turk (MTurk), which is an online marketplace for work. Requesters (employers) pay providers (workers) a fee to perform a human intelligence task, or HIT. In recent years, MTurk has become a popular tool to gain research participants (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler & Ipeirotis, 2010).

This study included 145 participants, of which 57% were male and 69% were white. The average age was 35, and 76.7% held at least an Associate's degree. Participants were also asked if they had previous interviewing experience, of which 48.3% indicated they had no previous interviewing experience. However, 32.4% reported previously interviewing 1-10 candidates, 13.8% reported previously interviewing 11-30 candidates, and 5.5% reported previously interviewing over 30 candidates.

Participants were compensated \$0.50 to participate in this study. Data from participants that failed the attention checks or that finished the survey in less than 120 seconds were removed. This resulted in the removal of 11 cases.

### **Procedure and Design**

Participants were first presented an informational letter detailing risks, benefits, incentives, and a description of the study (Appendix D). Participants were told that the purpose of the study was to determine how well untrained raters are able to score job interviews. Participants were randomly assigned to either the low, intermediate, or high performance condition. For each performance condition, the same three structured interview questions were asked.

Participants listened to an audio recording for each of three interview items for a total of three audio recordings. The presentation of the questions was counterbalanced to ensure there were no ordering effects. Participants rated the candidate's response using the anchored rating scale that corresponds to the question (see Appendix B) immediately after listening to the corresponding item. After all three questions were rated, participants completed a demographic survey.

### **Materials**

The interview questions, anchored rating scale, and the scripted candidate answers of low, intermediate, and high performance associated with each question can be found in Appendix B. Candidate responses for each performance condition were designed to represent low, intermediate, and high performance.

### **Measures**

Participants were asked to use a behaviorally anchored rating scale for each interview question. Participants were also asked demographic information including their age, gender, ethnicity, employment status, occupation, and previous experience interviewing job candidates. All interview items and anchored rating scales used can be found in Appendix C.

## Results

### Data Screening

Prior to performing any analyses, data were screened for potential outliers. First, minimum and maximum scores were examined and all data was found to be within the range of the scale. Next, the data were standardized; no cases were found to have an unusually large  $z$ -score ( $\pm 3$ ). In addition, data was inspected visually using a P-P plot to ensure that the assumption of normality was met. Further, skewness and kurtosis values were within normal range of  $\pm 3$ .

### Data Analysis

First, the dependent variable, interview ratings, was computed using an average of participants' ratings of interview questions 1, 2, and 3. A one-way, independent ANOVA was used to determine if there were significant differences in interview ratings in the low, intermediate, and high performance conditions. Results indicated that there were significant differences between performance conditions,  $F(2, 144) = 58.06, p < .05, \omega^2 = .44$ . Post hoc Tukey tests indicate that each performance condition was significantly different from the other,  $p < .05$ .

The mean interview rating in the low performance condition ( $N = 50$ ) was 2.70 with a standard deviation of .83. The mean interview rating in the intermediate performance condition ( $N = 48$ ) was 3.25 with a standard deviation of .66. The mean interview rating in the high performance condition ( $N = 47$ ) was 4.31 with a standard deviation of .72.



## **Discussion**

The results of Pilot Study 1 followed the expected trend that the low interview performance condition would be rated the lowest, followed by the intermediate interview performance condition, and the high interview performance condition. This provides evidence that participants are able to distinguish between low, intermediate, and high levels of interview performance. As expected, interview ratings were highest in the high interview performance condition and lowest in the low interview performance condition. Interestingly, the mean interview ratings in the low performance condition was 2.7, when the candidate responses to each question in the low performance condition were designed to represent low performance, or a score of 1. This suggest that leniency error may be a problem for untrained raters.

## **CHAPTER THREE**

### **PILOT STUDY TWO**

Pilot Study Two served as a manipulation check of the cosmetics variable and to inform which cosmetics conditions to use in study three. In Pilot Study Two, participants watched a short introductory video of a candidate that is wearing one of five facial cosmetics applications ranging from no cosmetics to a very heavy, dark application of facial cosmetics. This was done to ensure that participants perceived differences between the cosmetics conditions as different applications of facial cosmetics can be subtle and difficult to identify. Likewise, Pilot Study Two also examined perceptions of makeup application quality to ensure that there were not significant differences in the perception of makeup application quality between conditions.

#### **Method**

##### **Participants**

Participants were recruited using Amazon's Mechanical Turk (MTurk). This study included 251 participants, of which 60% were male and 66% were white. The average age was 36, and 73.3% held at least an Associate's degree. Participants were also asked if they had previous interviewing experience, of which 45.4% indicated they had no previous interviewing experience. However, 37.5% reported previously interviewing 1-10 candidates, 7.6% reported previously interviewing 11-30 candidates, and 9.6% reported previously interviewing over 30 candidates. Participants were compensated \$.50

to participate in this study. Data from 4 participants that failed attention checks were removed.

### **Procedure and Design**

Participants were first presented an information letter detailing risks, benefits, incentives, and a description of the study (Appendix D). Participants were told that the purpose of this research was to explore the first impressions of untrained raters to candidates interviewing for a job.

Next, participants viewed a short introductory video of a candidate interviewing for the role of Human Resources Manager. The candidate in the video was a 29 year old, white female with a body mass index in the normal range. In the video she said, “Hi, my name is Emily Howard and today I’ll be interviewing for the role of Human Resource Manager.” The job of Human Resources Manager was chosen in order to make the sex-type of the job gender-neutral.

Participants were randomly assigned to one of five cosmetics conditions. Depending on the cosmetics condition to which the participant was assigned, the candidate in the video seen by participants wore a different amount of facial cosmetics. Facial cosmetics ranged from wearing no facial cosmetics (condition one) to wearing a very heavy, glamorous application of facial cosmetics (condition five).

After participants watched the video, they then completed the survey measures. After participants completed the survey measures, they were debriefed (Appendix D) that the true purpose of the study was to examine how the use of facial cosmetics influences perceptions of job candidates.

## **Materials**

In condition one, the candidate did not wear any facial cosmetics. In condition two, the candidate wore foundation, mascara, and a light application of brow pencil. In condition three, the candidate wore foundation, blush, a light application of brow pencil, mascara, eyeliner, a light application of eyeshadow, and tinted lip color. In condition four, the candidate wore foundation, blush, contouring powder, a heavy application of brow pencil, mascara, winged eyeliner, a heavier application of eyeshadow, and lipstick. In condition five, the candidate wore foundation, blush, contouring powder, highlighter, a heavy application of brow pencil, mascara, winged eyeliner, a very heavy application of eyeshadow, lipstick, and false eyelashes. These cosmetics conditions were designed to range from a very light, natural makeup look (condition two), to a very heavy, glamorous makeup look (condition five). Photos of each cosmetics condition can be found in Appendix A. The URL for each video can be found in Appendix E.

## **Measures**

All items used can be found in Appendix C.

**Demographics.** Participants were asked their age, gender, ethnicity, employment status, occupation, and previous experience interviewing job candidates.

**Makeup Amount.** Participants were asked two items related to the amount of makeup the candidate wore in the video. The first item was “How much makeup would you say the candidate was wearing?” and used a rating scale of 1 (No makeup at all) to 5 (A great deal of makeup). The second item was “It was obvious that the candidate was wearing makeup” and used a rating scale of 1 (strongly disagree) to 5 (strongly agree).

**Makeup Application Quality.** Makeup application quality was measured using the item “The candidate’s makeup was applied well” on a scale of 1 (strongly disagree) to 5 (strongly agree).

**Attention checks.** In order to ensure that participants were actively participating in the study and responding intentionally, two attention checks were used. For example, participants were asked to respond “agree” or “disagree” to an item.

## Results

### Data Screening

Prior to performing any analyses, data were screened for potential outliers. First, minimum and maximum scores were examined and all data was found to be within the range of the scale. Next, the data were standardized; no cases were found to have an unusually large  $z$ -score ( $\pm 3$ ). Data was also inspected visually using a P-P plot to ensure that the assumption of normality was met. Further, skewness and kurtosis values were within normal range of  $\pm 3$ .

### Data Analysis

First, a one-way ANOVA was used to determine if there were perceived differences in the quality of makeup application between the cosmetics conditions. Data from the first condition were removed because the candidate was not wearing cosmetics in that condition. Results indicated that there were no significant differences in perceptions of makeup application quality between groups,  $F(3, 196) = 2.50, p = .06$ . The mean rating across conditions was 4.35, with a standard deviation of .84. Descriptive statistics for each condition can be found in Table 1.

Next, two additional one-way ANOVAs were used to examine perceptions of the amount of makeup worn in each condition. These perceptions were measured using two items: “It was obvious that the candidate was wearing makeup” and “How much makeup would you say the candidate was wearing?” Means for both items followed the expected trend, with the lowest scores in Condition 1 (no makeup) and rising slightly through each condition. Descriptive statistics can be found in Tables 2 and 3, respectively.

Using the item “How much makeup would you say the candidate was wearing?” as the dependent variable, results of the one-way ANOVA indicated that there were significant differences between cosmetics conditions,  $F(4, 246) = 16.02, p < .05, \omega^2 = .19$ . Post hoc Tukey tests indicated that Condition 1 was significantly different from Conditions 3, 4, and 5,  $p < .05$ ; Condition 2 was significantly different from Condition 5,  $p < .05$ ; Condition 3 was significantly different than Conditions 1 and 5,  $p < .05$ ; Condition 4 was significantly different than Condition 1,  $p < .05$ ; and Condition 5 was significantly different than Conditions 1, 2, and 3,  $p < .05$ .

The next dependent variable examined was “It was obvious that the candidate was wearing makeup.” Levene’s test for homogeneity of variances indicated that there were significant differences in the variances between cosmetics conditions,  $F(4, 246) = 3.42, p < .05$ . Because the assumption of homogeneity of variances was violated, the results of the one-way ANOVA are presented using Welch’s F-ratio (Welch, 1951). Results indicated that there were significant differences between cosmetics conditions,  $F(4, 122.57) = 14.54, p < .05, \omega^2 = .18$ . Games-Howell post hoc tests were used as they are appropriate for data that does not meet the assumption of homogeneity of variances (Games & Howell, 1976). Post-hoc tests indicated that Condition 1 was significantly different than Conditions 4 and 5,  $p < .05$ ; Condition 2 was significantly different than Conditions 4 and 5,  $p < .05$ ; Condition 3 was significantly different than Condition 5,  $p < .05$ , Condition 4 was significantly different than Conditions 1 and 2,  $p < .05$ ; and Condition 5 was significantly different than Conditions 1, 2, and 3,  $p < .05$ . Although

Condition 1 and 3 were not significantly different at the .05 level,  $p$  was equal to .06, which is near significance.



## **Discussion**

The purpose of Pilot Study Two was to ensure that the cosmetics conditions used to test the hypotheses were perceived as different from one another. Results of this study indicate that Conditions 1, 3, and 5 were perceived as significantly different from one another. Specifically, results demonstrate that Condition 1 was perceived as less makeup than Conditions 3 and 5, that Condition 3 was perceived as more makeup than Condition 1 but less than Condition 5, and that Condition 5 was perceived as more makeup than Condition 3. Therefore, Study 3 will use cosmetics conditions 1, 3, and 5 to test the remaining study hypotheses. This ensures that each condition is perceived as meaningfully different from one another.

Additionally, Pilot Study 2 also sought to ensure that perceptions of cosmetic application quality was high. This was to ensure that poor application quality did not bias or impede hypothesis testing. Results indicated that participants in each cosmetics condition did not significantly differ in their perceptions of makeup application quality. Further the mean rating across conditions was 4.35, which indicates perceptions of makeup application quality were generally high.

## **CHAPTER FOUR**

### **STUDY THREE**

The purpose of study three was to examine the effects of facial cosmetics on structured interview scores. Study three also examined how facial cosmetics relate to attractiveness, professional appearance, and perceived competence and if any of these variables mediate the relationship between facial cosmetics and structured interview scores. In addition, study three also examined interview performance as a potential moderator of the cosmetics-interview performance relationship.

#### **Method**

##### **Participants**

Participants were recruited using Amazon's Mechanical Turk (MTurk). This study included 452 participants, of which 56% were male and 75% were white. The average age was 37, and 68.1% held at least an Associate's degree. Participants were also asked if they had previous interviewing experience, of which 47.3% indicated they had no previous interviewing experience. However, 31.6% reported previously interviewing 1-10 candidates, 9.3% reported previously interviewing 11-30 candidates, and 11.7% reported previously interviewing over 30 candidates. Participants were compensated \$1.00 to participate in this study. Data from participants that failed the attention checks or that finished the survey in less than 120 seconds were removed. This resulted in the removal of 21 cases.

##### **Power Analysis**

A power analysis was performed to ensure that there was sufficient power to find the proposed effects. Using the program G\*Power, several power analyses were performed. First, a power analysis was performed to determine the number of participants needed to find a medium-sized effect for Hypotheses 1a, 1b, 2a, 2c, and 3a. Results indicated that to achieve .80 power, a sample of 159 participants was required. Next, a power analysis was performed to determine the number of participants needed to find a small effect for Hypothesis 1c, 2b, and 3b. Results indicated that to achieve .80 power, a sample of 485 was required.

### **Procedure and Design**

This study utilized a three (low performance, intermediate performance, high performance) by three (no cosmetics, low cosmetics, high cosmetics) between-subjects design. Participants were first presented an information letter detailing risks, benefits, incentives, and a description of the study (Appendix D). Participants were then asked to watch three videos. Participants were randomly assigned to a cosmetics and a performance condition, which influenced which specific videos they were shown. Each video contained one of the three interview items. The order in which each video was presented was counterbalanced to avoid ordering effects.

After participants watched each video, they were asked to rate the candidate's response using the behaviorally anchored rating scale that corresponded to the question (see Appendix B). After all three videos were watched and questions were rated, the participants completed the remaining survey measures. After participants completed the

survey measures, they were debriefed (Appendix D) that the true purpose of the study was to determine how well untrained raters are able to score job interviews.

## **Materials**

The candidate in the interview videos was held constant for all conditions in order to control for attractiveness and vocal cues such as articulation and pitch. Each answer was spoken at the same rate. The candidate was a 29-year-old, white female with a body mass index in the normal range. She had straight brunette hair that was worn down. The videos showed the candidate sitting, from the waste up. The candidate wore a neutral colored ivory blouse, a black suit jacket, and pearl earrings. Photos can be seen in Appendix A.

In the videos, a male voice read the interview item off screen while the camera focused on the candidate. The candidate then gave her answer. A total of 27 videos were created (3 cosmetics conditions x 3 performance levels x 3 interview items) and they ranged from 34 to 45 seconds in length. The URL for each video can be found in Appendix E.

Results of Pilot Study 2 indicated that cosmetics conditions 1 (no cosmetics), 3, and 5 were meaningfully different from one another. Therefore, those are the conditions that were used in this study. However, condition 3 was renamed cosmetics condition 2 or low cosmetics, and condition 5 was renamed condition 3 or high cosmetics. In cosmetics condition 1 or no cosmetics condition, the candidate did not wear any facial cosmetics, except a moisturizer/sunscreen that does not have any color or tint. In cosmetics condition 2, the low cosmetics condition, the candidate wore what is described by Russell

(2010) as the “received style” of cosmetics which are those that are commonly used and accepted to even skin tone and darken the eyes and the mouth. The products used for this style of cosmetics are foundation, blush, eyeliner, mascara, eyeshadow, and lipstick. Eyebrow pencil was also used to lightly fill in the eyebrows, providing additional facial contrast. Lastly, in the high cosmetics condition, the candidate wore a heavier application of makeup including contouring powder, highlighter, a heavier application of eyeshadow, false eyelashes, and a darker lip color. In this condition, the cosmetics should be noticeable to the average person. Photos of each cosmetics condition with a description of the products used can be found in Appendix A. The interview questions, anchored rating scale, and the scripted candidate answers of low, intermediate, and high performance answers associated with each question can be found in Appendix B.

## **Measures**

All items used can be found in Appendix C.

**Demographics.** Participants were asked their age, gender, ethnicity, employment status, occupation, and previous experience interviewing job candidates.

**Physical attractiveness.** Physical attractiveness was measured using the following item, “How physically attractive did you find the candidate?” using a rating scale of 1 (Very Unattractive) to 5 (Very Attractive).

**Professional appearance.** Professional appearance was measured using 7 items with a rating scale of 1 (Strongly Disagree) to 5 (Strongly Agree). An example item is, “The candidate put effort into looking professional.” Cronbach’s alpha for this scale was .87.

**Perceived Competence.** Perceived competence was measured using the following item, “How competent do you think the candidate is for the job?” using a rating scale of 1 (Not at all competent) to 5 (Very Competent).

**Attention checks.** In order to ensure that participants were actively participating in the study and responding intentionally, two attention check items were included in the survey. An example attention check is “Please respond ‘agree’ for this item.” Data from participants that did not answer the attention checks correctly was excluded from all analyses.

## **Results**

### **Data Screening**

Prior to performing any analyses, data were screened for potential outliers. First, minimum and maximum scores were examined and all data was found to be within the range of the scale. Next, the data was standardized and all z-scores were found to be within the normal range of  $\pm 3$ . In addition, data was inspected visually using a P-P plot to ensure that the assumption of normality was met. Skewness and kurtosis values were within the normal range of  $\pm 3$ . To check the assumption of homogeneity of variances, the standardized residuals were plotted against the standardized predicted values. This was done to graphically check for violations of the assumption of homoscedasticity. The scatter plots revealed residuals that were relatively constant across all levels of predicted values.

To screen for multivariate outliers for Hypotheses 1c, 2b, and 3b Mahalanobis Distance and Cook's D were examined for each set of variables to ensure that no cases had undue leverage or influence. Tests for Hypothesis 1c, 2b, and 3b, excluded cases with Mahalanobis Distance values above the critical value of 13.82. Testing for Hypothesis 1c excluded 2 cases, testing for Hypothesis 2b excluded 6 cases, and testing for Hypothesis 3b excluded 9 cases. Cook's D values were within the normal range (less than 1).

### **Data Analysis**

Data was analyzed using SPSS version 23 and the version 2.16 of the PROCESS macro written by Andrew Hayes. Prior to analyses, descriptive statistics and correlations

between all continuous variables were examined (Tables 4-5). Significant, positive correlations were seen between all study variables.

Hypothesis 1a was tested using a one-way ANOVA to determine if ratings of physical attractiveness were higher in the cosmetics conditions (low and high cosmetics conditions) than in the no-cosmetics condition. Results indicated that there were significant differences in ratings of physical attractiveness between cosmetics conditions  $F(2, 449) = 4.44, p < .05, \omega^2 < .01$ . Post hoc Tukey tests indicated that the no cosmetics condition ( $M = 3.60, SD = .82$ ) was significantly different from the low cosmetics conditions ( $M = 3.83, SD = .82$ ) and the high cosmetics condition ( $M = 3.85, SD = .78$ ),  $p < .05$ , but the low and high cosmetics conditions were not significantly different from each other,  $p = .98$ . These results provide support for Hypothesis 1a; ratings of attractiveness were higher in the cosmetics conditions than in the no cosmetics condition.

Hypothesis 1b was tested using a one-way ANOVA to determine if interview ratings were higher in the cosmetics conditions (conditions 2 and 3) than in the no cosmetics condition (condition 1). The variable interview ratings was created by averaging participant ratings of interview questions 1, 2, and 3. Results indicated that there were no significant differences in interview scores between cosmetics conditions,  $F(2, 449) = .64, p = .53$ . These results do not support Hypothesis 1b; Interview ratings were not significantly different in any of the cosmetics conditions. Means and standard deviations for each condition can be seen in Table 6.

Hypothesis 1c proposed that physical attractiveness would mediate the relationship between cosmetics and interview ratings. To test this hypothesis, the



mediation procedures described by Hayes and Preacher (2014) were used using Version 2.16 of the PROCESS macro written by Andrew Hayes (see Figure 3 for the mediation model). The independent variable was cosmetics condition (dummy coded such that the no cosmetics condition was the reference group), the mediator was ratings of attractiveness, and the dependent variable was interview ratings.

When the independent variable is multicategorical, there is not a significance test that can be interpreted (e.g. Sobel test). Rather, multiple ( $k-1$ , where  $k$  is the number of IV categories) parameter estimates are examined. Specifically, Hayes and Preacher (2014) recommend examining what they call the “relative indirect effects” which are the indirect effects of each dummy variable on the dependent variable through the mediator. The relative indirect effect of cosmetics dummy variable 1 (coded to represent the low cosmetics condition) on interview ratings through attractiveness is labeled “ $a_1b$ ” and the relative indirect of cosmetics dummy variable 2 (coded to represent the high cosmetics condition) on interview ratings through attractiveness is labeled “ $a_{1b}$ .” If any one of the relative indirect effects ( $a_1b$ ,  $a_{1b}$ ) is different from zero, then it can be concluded that mediation has occurred (Hayes & Preacher, 2014). Results indicated that both relative indirect effects were significantly different from 0,  $a_1b$   $b = .064$ , 95% BCa CI [.017, .137],  $a_{1b}$   $b = .069$ , 95% BCa CI [.023, .142]. Confidence intervals are based on 5000 bias corrected bootstrap samples. These results indicate that as there is a one unit change in cosmetics (i.e. from condition 1 to 2 or from 1 to 3), we can expect interview ratings to increase by .064 and .069, respectively, through the effect of cosmetics on attractiveness, which then influences interview ratings.

These relative indirect effects were standardized so that the coefficients could be more easily compared. This was done using the formula:  $(a_{k-1}b/SD_{\text{outcome}})*SD_{\text{predictor}}$ , where  $a_{k-1}b$  is the respective relative indirect effect. The standardized relative indirect effect for the low cosmetics condition = .05, and the standardized relative indirect effect for the high cosmetics condition = .06. These results provide some support for Hypothesis 1c; cosmetics have a small, indirect effect on interview ratings through physical attractiveness. A diagram of the full mediation model, including all relative direct and relative indirect effects can be seen in Figure 4.

Hypotheses 2a and 2c were tested using a one-way ANOVA to determine if ratings of professional appearance were higher in the cosmetics conditions (conditions 2 and 3) than in the no cosmetics condition (condition 1). Results indicated that there was a significant difference in ratings of professional appearance between cosmetics conditions,  $F(2, 441) = 3.56, p < .05, \omega^2 = .02$ . Post-hoc Tukey tests indicated that ratings of professional appearance in the low cosmetics condition ( $M = 4.74, SD = .44$ ) were significantly higher than ratings of professional appearance in the no cosmetics condition ( $M = 4.60, SD = .48$ ) and the high cosmetics condition ( $M = 4.66, SD = .41$ ),  $p < .05$ . There were not significant differences in ratings of professional appearance between the no cosmetics and high cosmetics conditions,  $p = .48$ . These results provide support for Hypothesis 2c as ratings of professional appearance were highest in the low cosmetics condition.

Hypothesis 2b proposed that professional appearance would mediate the relationship between cosmetics and interview ratings. This hypothesis was tested using

the Hayes and Preacher (2014) mediation procedures described above for Hypothesis 1c. The independent variable was cosmetics condition (dummy coded such that the no cosmetics condition was the reference group), the mediator was professional appearance, and the dependent variable was interview ratings. The relative indirect effect of cosmetics on interview ratings through professional appearance was significant for the low cosmetics condition as compared to the reference group,  $a_1b = .020$ , 95% BCa CI [.001, .061], but not significant for the high cosmetics condition as compared to the reference group,  $a_1b = .010$ , 95% BCa CI [-.001, .008]. Confidence intervals are based on 5000 bias corrected bootstrap samples. The standardized relative indirect effect of  $a_1b$  equals .01. These results indicate that there is evidence that the relationship between cosmetics and interview ratings is mediated by professional appearance. Specifically, as there is a one unit change in cosmetics (i.e. from condition 1 to 2), we can expect interview ratings to increase by .02, through the effect of cosmetics on professional appearance, which then influences interview ratings. Therefore cosmetics have a small indirect effect on interview ratings through professional appearance, which provides support for Hypothesis 2b. A diagram of the full mediation model, including all coefficients can be seen in Figure 5.

Hypothesis 3a was tested using a one-way ANOVA to determine if ratings of perceived competence were higher in the cosmetics conditions (conditions 2 and 3) than in the no cosmetics condition (condition 1). Results indicated that there were no significant differences in ratings of competence between cosmetics conditions,  $F(2, 449)$

= .05,  $p = .95$ . Therefore the data does not support Hypothesis 3a; ratings of perceived competence did not differ in any of the cosmetics conditions.

Hypotheses 3b proposed that perceived competence would mediate the relationship between cosmetics and interview ratings. This hypothesis was tested using the Hayes and Preacher (2014) mediation procedures described above for Hypothesis 1c. There was no evidence that the relationship between cosmetics and interview ratings is mediated by perceived competence, low cosmetics condition  $b = -.137$ , 95% BCa CI [- .329, .054], high cosmetics condition  $b = -.126$ , 95% BCa CI [-.318, .067]. These results indicate that perceived competence does not mediate the relationship between cosmetics and interview ratings. A diagram of the full mediation model, including all coefficients can be seen in Figure 6. Although there was no evidence of mediation, perceived competence did significantly predict interview ratings,  $b = .642$ ,  $p < .05$ , 95% BCa CI [.525, .759].

Research Question 1 sought to explore the potential moderating effect of employee performance on the relationship between facial cosmetics and interview ratings. A 3 (cosmetics) x 3 (performance) factorial ANOVA was used to determine if the effect of facial cosmetics is different at low, intermediate, and high levels of interviewee performance. Results indicated that there was a main effect of performance on interview ratings,  $F(2, 443) = 242.95$ ,  $p < .05$ ,  $\eta_p^2 = .52$ . Post-hoc Tukey tests revealed that interview ratings in performance condition 1 ( $M = 2.77$ ,  $SD = .77$ ) were significantly lower than interview ratings in performance condition 2 ( $M = 3.57$ ,  $SD = .59$ ) and in performance condition 3 ( $M = 4.42$ ,  $SD = .61$ ) at the  $p < .05$  level. Further, interview

ratings in performance condition 2 were significantly lower than in performance condition 3,  $p < .05$ .

Results indicated that there was not a main effect of cosmetics on interview ratings,  $F(2, 443) = 2.27, p = .11$ , and there was not a significant interaction between performance and cosmetics on interview ratings,  $F(4, 443) = 2.35, p = .05$ . Descriptive statistics can be seen in Table 7. Although the interaction was not significant, the interaction approached significance and results were graphed (Figure 7) and simple effects were examined. Simple effects indicated that at performance level 1, there was a significant difference in interview ratings between the no cosmetics condition ( $M = 3.00, SD = .76$ ) and the low cosmetics ( $M = 2.62, SD = .78$ ) and high cosmetics conditions ( $M = 2.66, SD = .72$ ). This indicates that at low levels of performance, interview ratings were higher in the no cosmetics condition than in the cosmetics conditions.

Lastly, because cosmetics was not a predictor of interview ratings but attractiveness was a predictor of interview ratings (see Figure 4), an analysis was performed to determine if the relationship between attractiveness and interview ratings may be moderated by performance. This was done using version 2.16 of the PROCESS macro written by Andrew Hayes. However, results indicated that performance does not moderate the relationship between attractiveness and interview ratings, attractiveness x performance condition 2 dummy variable  $b = -.18, p = .05, 95\% \text{ BCa CI } [-.36, .00]$ , attractiveness x performance condition 3 dummy variable  $b = -.15, p = .13, 95\% \text{ BCa CI } [-.34, .04]$ . Therefore, there is no evidence that performance moderates the relationship between attractiveness and interview ratings.

## **Discussion**

The purpose of study three was to examine the effects of facial cosmetics on structured interview ratings (Hypothesis 1b). Study three also sought to examine how facial cosmetics relate to attractiveness (Hypothesis 1a), professional appearance (Hypothesis 2a), and perceived competence (Hypothesis 3a) and if any of these variables mediate the relationship between facial cosmetics and structured interview ratings (Hypotheses 1c, 2b, and 3b). Lastly, study three examined interview performance as a potential moderator of the cosmetics-interview performance relationship.

Hypothesis 1a proposed that ratings of interviewee physical attractiveness would be higher in the cosmetics conditions than in the no-cosmetics condition. Results indicated that ratings of physical attractiveness were significantly higher in cosmetics conditions 2 and 3 than in cosmetics condition 1. This provides support for Hypothesis 1a; ratings of attractiveness were higher in the cosmetics conditions than in the no cosmetics condition. This result is consistent with previous research (Cash, Dawson, Davis, Bowen, & Galumbeck, 1989; Cox & Glick, 1986; Etcoff, Stock, Haley, Vickery, & House, 2011; Miller & Cox, 1982; Mulhern, Fieldman, Hussey, Leveque, & Pineau, 2003; Workman & Johnson, 1991).

Hypothesis 1b proposed that interview ratings would be higher in the cosmetics conditions than in the no cosmetics condition. Results indicated that there were no significant differences in interview ratings between cosmetics conditions. Therefore, no support was found for Hypothesis 1b.

Hypothesis 1c proposed that physical attractiveness would mediate the relationship between cosmetics and interview ratings. Results indicated that the relative indirect effects of cosmetics had a non-zero effect on interview ratings. Therefore, there is evidence of mediation, which supports Hypothesis 1c. This means that cosmetics have a small, indirect effect on interview scores through their effect on perceptions of physical attractiveness.

Hypothesis 2a proposed that ratings of interviewee professional appearance would be higher in the cosmetics conditions than in the no cosmetics condition. Results indicated that there were significant differences in ratings of professional appearance between cosmetics conditions, providing support for Hypothesis 2a. Hypothesis 2c proposed that ratings of professional appearance would be highest in the low cosmetics condition, followed by the high cosmetics condition, and lowest in the no-cosmetics condition. Results indicated that ratings of professional appearance were significantly higher in the low cosmetics condition than in either the no-cosmetics or high cosmetics conditions, which provides support for Hypothesis 2c. This result is consistent with previous research that also found differential effects on perceptions as a result of varying degrees or amounts of cosmetics (Etcoff et al., 2011), while the decrease in perceived professional appearance from low cosmetics to high cosmetics is consistent with previous research that there may be a penalty associated with wearing heavy makeup (Huguet et al, 2004).

Hypothesis 2b proposed that ratings of interviewee professional appearance would mediate the relationship between cosmetics and interview ratings. Results

indicated that the relative indirect effect of cosmetics condition 2 through professional appearance ( $a_1b$ ) had a non-zero effect on interview ratings. This indicates that cosmetics have a small indirect effect on interview ratings through professional appearance, which provides support for Hypothesis 2b.

Hypothesis 3a proposed that ratings of perceived interviewee competence would be higher in the cosmetics conditions than in the no cosmetics condition. Results indicated that there were no significant differences between cosmetics conditions, indicating that there is not support for Hypothesis 3a. These results are not consistent with previous research (Etcoff et al., 2011) that displayed photos of women wearing various amounts of facial cosmetics. These disparate findings may be attributable to the differences in stimuli medium (e.g., photos vs. videos). Hypothesis 3b proposed that perceived interviewee competence would mediate the relationship between cosmetics and interview ratings. Results indicated that the relative indirect effects did not have a significant effect on interview ratings, demonstrating that there is not support for hypothesis 3b.

Research Question 1 sought to explore the potential moderating effect of employee performance on the relationship between facial cosmetics and interview ratings. Previous cosmetics research has not examined the impact of performance on perceptions (Etcoff et al., 2011; Mileva et al., 2016; Nash et al., 2006), so it was unclear if cosmetics would be equally influential across different levels of performance or if there would be differential effects. Results did not indicate that there was any meaningful interaction between cosmetics and interview performance on interview ratings.



Therefore, the influence of facial cosmetics does not vary at different levels of interview performance. This is not surprising given that a direct effect of cosmetics on interview ratings was not found. However, a direct effect of attractiveness on interview scores was found. To further explore this research question, the effect of performance on the relationship between attractiveness and interview scores was examined as the mediation analysis from Hypothesis 1c indicated that attractiveness does have a direct effect on interview ratings. However, the moderation analysis indicated that the relationship between attractiveness and interview ratings does not vary at different levels of interview performance.

## **CHAPTER FIVE**

### **GENERAL DISCUSSION**

Popular media espouses the notion that it is necessary to wear facial cosmetics to have interview and career success (Cardellino, 2013; Khazan, 2015), however, until now, there has been no empirical research to support or deny this widespread claim. Although there was no previous research on the link between cosmetics and interview ratings, previous research has provided ample evidence that the use of facial cosmetics positively influences perceptions of attractiveness (Cox & Glick, 1986; Etkoff et al, 2011; Mulhern et al, 2003; Workman & Johnson, 1991). In turn, attractiveness positively influences job-related outcomes such as hiring decisions, performance evaluations (Hosoda et al., 2003), and interview ratings (Barrick et al., 2009). Therefore, the primary purpose of this dissertation was to explore the relationship between facial cosmetics and interview ratings.

I found support for several of my hypotheses. Overall, some evidence was found that facial cosmetics influence structured interview ratings. Specifically, cosmetics were found to influence interview ratings through the mediating variables of physical attractiveness and professional appearance. No direct link between cosmetics and interview ratings was present. These findings suggests that women should consider using facial cosmetics in preparation for an employment interview as it may improve their attractiveness and professional appearance, which in turn improves interview ratings. However, it is important to note that the relationships found in this study were very small,

so whether or not cosmetics have meaningful influence on post-interview decision making is yet to be seen.

In addition to the indirect effect of cosmetics on interview ratings through physical attractiveness and professional appearance, the mediation analyses demonstrated that both attractiveness and professional appearance predict interview ratings. These results are consistent with previous research (Barrick et al., 2009) and have practical implications for those affected by employment interviews. First, the effort made by job candidates to increase their attractiveness and present themselves professionally is a worthwhile investment that may result in higher interview scores. Additional research is necessary to determine what appearance practices result in an optimally professional appearance as it is likely that different individuals hold a range of opinions on what is and is not considered to be professional. Second, organizations using interviews as a part of their selection system should take steps to minimize this bias in their employment interviews. Although making an effort to be perceived as more attractive or professional is advantageous to candidates, this is a source of error in interview ratings that could influence an organization's ability to select the best candidates. As demonstrated by previous research, organizations could minimize biased ratings through interviewer training and/or increased interview structure (Huffcutt et al., 2013; Lin et al., 1992).

Consistent with previous research, this study also found that ratings of physical attractiveness were higher in the cosmetics conditions than in the no cosmetics condition. Expanding upon previous work, this is the first study to link the use of facial cosmetics to ratings of professional appearance. Specifically, ratings of professional appearance were

higher in cosmetics condition 2, which included foundation, blush, eyebrow pencil, mascara, eyeliner, eyeshadow, and lip color applied in a light, professional manner. This indicates that the candidate was perceived less professionally when no facial cosmetics were worn and when a very heavy application of facial cosmetics were worn. These results provide some evidence that the amount and/or type of facial cosmetics worn must be considered as not all applications of cosmetics will be perceived equally. Future research should further explore this effect as the type or amount of cosmetics considered professional or appropriate likely varies based on the context of the job and/or organization. For example, the facial cosmetics considered professional for an attorney are likely to be very different than the facial cosmetics considered professional for a musician. The specific culture of an organization may also influence perceptions of what types or amounts of cosmetics are considered professional.

This study failed to find evidence that facial cosmetics influence perceptions of competence, which is inconsistent with previous findings (Etcoff et al., 2011). It is possible that this relationship was not found because there are other missing mediators of the relationship. For example, similar to the relationship between facial cosmetics and interview ratings, the relationship between facial cosmetics and perceived competence may be mediated through other variables, such as physical attractiveness or professional appearance. It is also possible that facial cosmetics do not influence perceptions of competence within the individual. The previous study that demonstrated a link between cosmetics and perceptions of competence used 25 different models wearing different amounts of makeup, whereas the current study only used 1 model. Therefore, it is

possible that the effects are so small they do not emerge with just one individual. Lastly, it is also possibly that because performance information was available to raters, there was less need to rely on initial impressions or biases when ratings the candidate's competence.

Lastly, this study also examined the potential moderating effect of interview performance on both the cosmetics-interview ratings relationship and the attractiveness-interview ratings relationship. However, there was no evidence that performance was a significant moderator of either relationship. This suggest that regardless of interview performance, the biasing effect of attractiveness on interview ratings remains constant. These results imply that the biasing influence of attractiveness is a problem for candidates interviewing for a job, regardless of their skill or competence.

### **Limitations**

The first limitation of this study is that the cosmetics stimuli was manipulated via video opposed to in-person interviews. It is likely that participants were not able to see the full detail and extent of the cosmetics via video. Therefore, it is possible that perceptions of and reactions to the candidate's cosmetics could be different when observed in-person. This is especially relevant since many interviews are held in-person, opposed to over videoconference. However, it is important to note that the use of video is an improvement over previous studies that used only photos to examine the influence of cosmetics on perceptions (Etcoff et al, 2011; Nash et al., 2006).

Another limitation of this study is that the same candidate was used in all conditions, which limits the conclusions that can be drawn by this study. For example,

because only one candidate was used in this study, only perceptions of attractiveness were varied, not actual physical attractiveness. Therefore the conclusions drawn by this study are limited to candidates of a similar level of attractiveness, age, and race. Further, it is possible that facial cosmetics improve the attractiveness of some individuals more or less than others. Future research should address these limitations by repeating this experiment with women of different ages, ethnicities, and physical attractiveness levels. This would further inform the relationship between facial cosmetic and interview ratings.

Lastly, the participants in this study were not trained interviewers, nor were they provided training as part of the study. Training interviewers is a factor identified by Campion et al. (1997) as improving the evaluation of interviews. This study found that participants were able to distinguish between the low, intermediate, and high levels of interview performance, but the mean ratings in each of these conditions suggest that the central tendency rating error and leniency error were likely problems for many of the participants. The interview answers given were designed to represent scores of 1, 3, and 5 on the behaviorally anchored ratings scale. Nevertheless, the actual mean interview ratings in Study 3 were 2.77, 3.57, and 4.42. This indicates that a large amount of variance in the ratings is likely due to error, which makes it more difficult to evaluate the true relationships between variables. Previous research has demonstrated that by training evaluators, rating errors can be reduced (Ivancevich, 1979; Pulakos, 1984).

## **Conclusion**

This dissertation is the first foray into examining the influence of facial cosmetics on interview ratings and provides several interesting contributions as well as new

pathways to be explored. Overall, this study found that facial cosmetics do not directly influence interview ratings but do have a small positive effect on interview ratings through the mediating influence of physical attractiveness and professional appearance. These results provide some small support for the common advice that it is important to wear makeup to job interviews. However, effect sizes were very small. In addition, this study was the first to link facial cosmetics to ratings of professional appearance. Interestingly, ratings of professional appearance were significantly higher in the low cosmetics condition than in the no cosmetics and high cosmetics conditions, suggesting that the amount of makeup worn has differential effects on perceptions of professional appearance. Specifically, this suggests that in order to be perceived as optimally professional, women should wear some makeup to job interviews, but not too much. Future research should continue to investigate the relationship between cosmetics and interview ratings with women of different ages, ethnicities, and attractiveness levels to better understand the effect.

## APPENDICES



## Appendix A

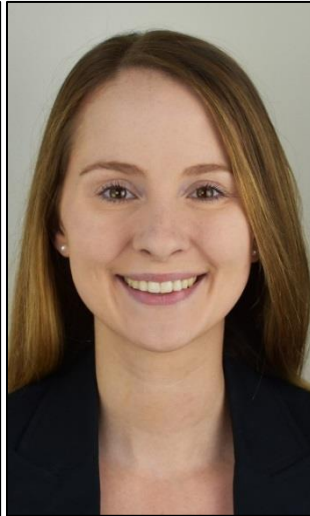
### Cosmetics Conditions Presented in Pilot Study Two and Study Three

#### Pilot Study Two Cosmetics Conditions

Condition 1



Condition 2



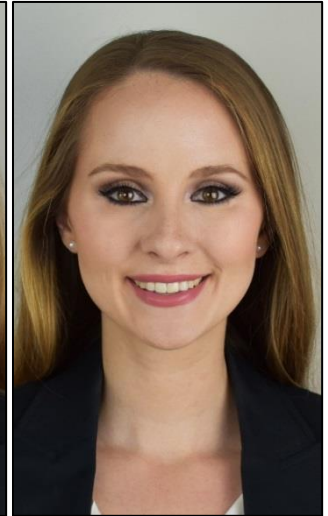
Condition 3



Condition 4



Condition 5



Condition 1 – The candidate is not wearing any facial cosmetics.

Condition 2 – The candidate is wearing foundation, mascara, and a light application of brow pencil.

Condition 3 – The candidate is wearing foundation, blush, a light application of brow pencil, mascara, eyeliner, a light application of eyeshadow, and tinted lip color.

Condition 4 – The candidate is wearing foundation, blush, contouring powder, a heavy application of brow pencil, mascara, winged eyeliner, a heavier application of eyeshadow, and lipstick.

Condition 5 – The candidate is wearing foundation, blush, contouring powder, highlighter, a heavy application of brow pencil, mascara, winged eyeliner, a very heavy application of eyeshadow, lipstick, and false eyelashes.

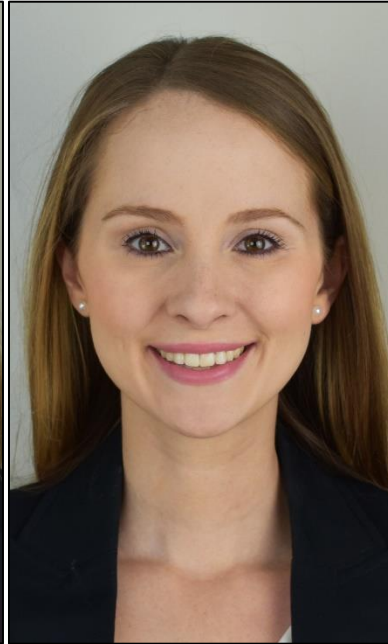
## Appendix A (continued)

### Study Three Cosmetics Conditions

Condition 1 - No Cosmetics



Condition 2 -Low  
Cosmetics



Condition 3 - High  
Cosmetics



Condition 1 (No Cosmetics) – The candidate is not wearing any facial cosmetics.

Condition 2 (Low Cosmetics) – The candidate is wearing foundation, blush, a light application of brow pencil, mascara, eyeliner, a light application of eyeshadow, and tinted lip color.

Condition 3 (High Cosmetics) – The candidate is wearing foundation, blush, contouring powder, highlighter, a heavy application of brow pencil, mascara, winged eyeliner, a very heavy application of eyeshadow, lipstick, and false eyelashes.

## Appendix B

### Interview Questions, Rating Scales, and Candidate Responses

*Question 1 - You are presenting a new project idea that you have invested a lot of time and effort into preparing. One of your colleagues immediately questions the utility of the project and then starts having a side conversation during your presentation. Please describe how you would behave in this situation.*

This item was adapted from an item used in by Ingold, Kleinmann, Konig, Melchers, and Van Iddekinge (2015).

Rating Scale	
5:	Firmly asks the colleague to refrain from his or her conversation, addresses skeptical arguments, and continues with the presentation.
4:	
3:	Bides his or her time, tries to ignore the conversation and asks the colleague to stop after quite some time using a moderate tone or gives him or her disapproving looks.
2:	
1:	Ignores the side conversation or gives in and breaks off the presentation.

**Low Performance Response** - “When someone’s talking during a presentation, it really reflects more poorly on them than it does on me. So if I was in that situation, I would do my best to continue the presentation and present my ideas, without letting that side conversation distract me or throw me off. I’d just focus on giving my presentation to the best of my ability. Hopefully there would still be other people paying attention and my presentation would still go well.”

**Intermediate Performance Response** – “That’s a tough situation because sometimes people don’t realize how loud or distracting their side conversations can be to the person presenting. But in this case, I would probably do my best to defend the usefulness of the

## Appendix B (continued)

project and then focus on giving the presentation to the best of my ability. If they didn't get the hint and the conversation continued, I would ask them to please end their conversation until I was finished."

**High Performance Response** – "First, I would try to address the concerns that my colleague brought up. If I've done my due diligence on the project, then I'll be able to give a detailed response why the project does have utility. Regarding the side conversation, I would ask that colleague to please end the conversation. I would make sure to be respectful, but also firm and direct. Then I would do my best to give the presentation to the best of my ability"

*Question 2 - The team you supervise works closely together on many projects. Often, the work of one person cannot be completed until the work of another is complete. An employee that you supervise is frequently missing deadlines, which then causes delays for the entire team. Please describe how you would behave in this situation.*

Rating Scale	
5:	Asks the employee why they're struggling to meet deadlines, offers help or suggestions for better time management. Maintains respectful, professional demeanor.
4:	
3:	Informs the employee that they have been missing deadlines, explains why that's a problem for the team.
2:	
1:	Admonishes the employee for missing deadlines, demands that future deadlines be met or jumps to punishment.

**Low Performance Response** – "This is one of those situations that no manager every really wants to be in. So in this case, I would have a very direct conversation with the employee about how missing deadlines is unacceptable. Everyone has expectations that they have to meet. I would let them know that going forward, I expect them to meet their

## Appendix B (continued)

deadlines, and if it continues to be a problem in the future, I'd have to explore disciplinary options. I don't like disciplining employees, but sometimes it's necessary."

**Intermediate Performance Response** – "I think the most important thing a supervisor can do in a situation like that is to just be honest. So I would start by saying something along the lines of 'Hey, you know, I've noticed that you've been struggling to make deadlines lately' and then let them know that when they miss deadlines, it has a direct effect on other people's ability to get work done. So ideally, that would be enough encouragement to keep them on track in the future."

**High Performance Response** – "I would start by meeting with the employee and letting them know that I've noticed that they're missing deadlines, and that has a negative impact on the rest of the team. I would ask them why they haven't been able to meet these deadlines...because if I can get to the root of the problem, I can help the employee solve the problem. For example, if they're struggling to prioritize their work or manage their time, I can give them guidance on how to improve."

*Question 3 - Please tell us about a time when you had to make a difficult decision in the past. How did you make your decision and what was the outcome?*

Rating Scale	
5:	Takes a strategic approach to analyzing information, weighs the pros and cons, considers different alternatives.
4:	
3:	Chooses an option that meets the requirements of the situation, doesn't consider whether their choice was the <i>best</i> option
2:	
1:	Makes their decisions based on "gut feelings" or instincts.

## **Appendix B (continued)**

**Low Performance Response** – “At my first job, I was responsible for planning a company lunch n’ learn on a health-related topic. My first instinct was that health-related topics had already been sufficiently covered in the past, so I decided to present on a new technology that the company had just purchased. I thought this was a better topic because I already had experience using the technology, and I’d heard my coworkers express some interest in how to use it. The lunch n’ learn ended up going really well and I got a lot of positive feedback from the attendees.”

**Intermediate Performance Response** – “At my first job, I was responsible for planning a company lunch n’ learn on any health-related topic. The company had already covered healthy cooking, stress management, and wellness myths. So I chose to present on in-office exercising to switch things up a bit. Plus, I was already familiar with the topic, so I knew I would be able to do a good job. It ended up going really well. I got a lot of positive feedback from the attendees and I even saw some of them doing the exercises that were taught in the class.”

**High Performance Response** - “I was responsible for planning a company lunch n’ learn on any health-related topic. I was new to the company, so I started by reviewing what topics had already been covered. Then, I made a list of new options and sent out a survey to determine which topics my coworkers were interested in. The most highly rated topics were Healthy Cooking and In-Office Exercising, but I chose In-Office Exercising because it was the more cost-effective option. I got a lot of positive feedback and I even saw some of people doing the exercises later.”

## Appendix C

### Measures

#### Demographics

What is your age?

What is your gender?

- Male
- Female
- Prefer not to respond

What is your ethnicity?

- White
- Black or African American
- Hispanic or Latino
- Asian
- Native American or American Indian
- Native Hawaiian or Pacific Islander
- Other

What is your highest level of education?

- No high school diploma or equivalent
- High school graduate, diploma or the equivalent (for example: GED)
- Some college credit, no degree
- Trade/technical/vocational training
- Associate degree
- Bachelor's degree
- Master's degree
- Professional degree
- Doctorate degree

What is your employment status?

- Employed full time
- Employed part time
- Not employed - looking for work
- Not employed - not looking for work
- Retired
- Student

If you are currently employed, what is your occupation?

Have you ever been responsible for interviewing job candidates?

- No
- Yes, interviewed 1-10 candidates in the past
- Yes, interviewed 11-30 candidates in the past
- Yes, interviewed over 30 candidates in the past

## **Appendix C (continued)**

### **Physical Attractiveness**

1. How physically attractive did you find the candidate?

### **Professional Appearance**

1. The candidate displayed a professional appearance.
2. The candidate put effort into looking professional.
3. The candidate was neatly groomed.
4. The candidate was dressed professionally.
5. The candidate was wearing an appropriate amount of makeup.
6. The candidate's hair was appropriately styled.
7. The candidate looked appropriate for a job interview.

### **Makeup Amount**

1. It was obvious the candidate was wearing makeup.
2. How much makeup would you say the candidate was wearing?

### **Makeup Application Quality**

1. The candidate's makeup was applied well.

### **Perceived Competence**

1. How competent do you think the candidate is for the job?

### **Perceived Intelligence**

1. How would you rate the candidate's intelligence?

### **Hiring Recommendation**

1. The candidate should be hired for the job.
2. I would recommend the candidate for hire.

### **Perceived Social Skills**

1. How would you rate the candidate's social skills?

### **Confidence in Ratings**

1. How confident are you that you scored question 1 correctly?
2. How confident are you that you scored question 2 correctly?
3. How confident are you that you scored question 3 correctly?



## **Appendix D**

### **Information Letters and Debriefings**

#### **Pilot Study 1 Information Letter (consent):**

Information about Being in a Research Study - Clemson University

#### **Description of the Study and Your Part in It**

Skye Gillispie and Robert Sinclair invite you to take part in a research study. Robert Sinclair is a professor at Clemson University. Skye Gillispie is a student at Clemson University, running this study with the help of Robert Sinclair. The purpose of this research is to determine how well untrained raters are able to score job interviews.

Your part in the study will be to listen to three audio recordings of a job interview and then complete a survey on your reactions to the recordings.

It will take you about 10 minutes to complete this study.

#### **Risks and Discomforts**

We do not know of any risks or discomforts to you in this research study.

#### **Possible Benefits**

We do not know of any way you would benefit directly from taking part in this study. However, this study will advance the academic community's knowledge and understanding of the ability of untrained raters to score job interviews.

#### **Incentives**

Upon successful completion of this study, including all attention checks, you will be compensated \$.50. If you do not successfully complete the attention checks, you will not be compensated.

#### **Protection of Privacy and Confidentiality**

We are not collecting any personally identifying information and will do everything we can to protect the confidentiality of the data. The results of this study may be published in scientific journals, professional publications, or educational presentations; however, no individual participant will be identified.

## **Appendix D (continued)**

### **Choosing to Be in the Study**

You do not have to be in this study. You may choose not to take part and you may choose to stop taking part at any time. You will not be punished in any way if you decide not to be in the study or to stop taking part in the study.

### **Contact Information**

If you have any questions or concerns about this study or if any problems arise, please contact Robert Sinclair at Clemson University at 864-656-393 or by email at [rsncla@clemson.edu](mailto:rsncla@clemson.edu).

If you have any questions or concerns about your rights in this research study, please contact the Clemson University Office of Research Compliance (ORC) at [irb@clemson.edu](mailto:irb@clemson.edu).

You may print a copy of this informational letter for your files.

### **Pilot Study 2 Information Letter (consent):**

Information about Being in a Research Study at Clemson University

#### **Description of the Study and Your Part in It**

Skye Gillispie and Robert Sinclair invite you to take part in a research study. Robert Sinclair is a professor at Clemson University. Skye Gillispie is a student at Clemson University, running this study with the help of Robert Sinclair. The purpose of this research is to explore the first impressions of untrained raters to candidates interviewing for a job.

Your part in the study will be to watch a short video of a candidate about to be interviewed for a job and then complete a survey on your reactions to the video.

It will take you 5-10 minutes to complete this study.

#### **Risks and Discomforts**

We do not know of any risks or discomforts to you in this research study.

#### **Possible Benefits**

## **Appendix D (continued)**

We do not know of any way you would benefit directly from taking part in this study. However, this study will advance the academic community's knowledge and understanding of how untrained raters react to job candidates.

### **Incentives**

Upon successful completion of this study, including all attention checks, you will be compensated \$.50. If you do not successfully complete the attention checks, you will not be compensated.

### **Protection of Privacy and Confidentiality**

We are not collecting any personally identifying information and will do everything we can to protect the confidentiality of the data. The results of this study may be published in scientific journals, professional publications, or educational presentations; however, no individual participant will be identified.

### **Choosing to Be in the Study**

You do not have to be in this study. You may choose not to take part and you may choose to stop taking part at any time. You will not be punished in any way if you decide not to be in the study or to stop taking part in the study.

### **Contact Information**

If you have any questions or concerns about this study or if any problems arise, please contact Robert Sinclair at Clemson University at 864-656-393 or by email at [rsncla@clemson.edu](mailto:rsncla@clemson.edu).

If you have any questions or concerns about your rights in this research study, please contact the Clemson University Office of Research Compliance (ORC) [irb@clemson.edu](mailto:irb@clemson.edu).

You may print a copy of this informational letter for your files.

### **Pilot Study 2 Debriefing:**

Additional Information:

Thank you for taking part in this study. You were told at the beginning of the study that the purpose of this study was to explore first impressions of untrained raters to candidates interviewing for a job. Now that you have completed the

## **Appendix D (continued)**

study, we want to let you know that the true purpose of this study was actually to examine how the use of facial cosmetics influences perceptions of job candidates. We did not tell you the true purpose of this study because we did not want you to focus on the candidate's appearance more or less than you naturally would.

If you would like a copy of the results of the study once it is completed, you may contact Skye Gillispie at [sgillis@clemson.edu](mailto:sgillis@clemson.edu). Thank you again for taking part in this study!

### **Study 3 Information Letter (consent):**

Information about Being in a Research Study - Clemson University

#### **Description of the Study and Your Part in It**

Skye Gillispie and Robert Sinclair invite you to take part in a research study. Robert Sinclair is a professor at Clemson University. Skye Gillispie is a student at Clemson University, running this study with the help of Robert Sinclair. The purpose of this research is to determine how well untrained raters are able to score job interviews.

Your part in the study will be to watch a three short videos of a candidate being interviewed for a job and then complete a survey on your reactions to the videos.

It will take you about 10-15 minutes to complete this study.

#### **Risks and Discomforts**

We do not know of any risks or discomforts to you in this research study.

#### **Possible Benefits**

We do not know of any way you would benefit directly from taking part in this study. However, this study will advance the academic community's knowledge and understanding of the ability of untrained raters to score job interviews.

#### **Incentives**

Upon successful completion of this study, including all attention checks, you will be compensated \$1.00. If you do not successfully complete the attention checks, you will not be compensated. Additionally, if you do not spend enough time completing the survey to watch all three videos, you will not be compensated.

## **Appendix D (continued)**

### **Protection of Privacy and Confidentiality**

We are not collecting any personally identifying information and will do everything we can to protect the confidentiality of the data. The results of this study may be published in scientific journals, professional publications, or educational presentations; however, no individual participant will be identified.

### **Choosing to Be in the Study**

You do not have to be in this study. You may choose not to take part and you may choose to stop taking part at any time. You will not be punished in any way if you decide not to be in the study or to stop taking part in the study.

### **Contact Information**

If you have any questions or concerns about this study or if any problems arise, please contact Robert Sinclair at Clemson University at 864-656-393 or by email at [rsncla@clemson.edu](mailto:rsncla@clemson.edu).

If you have any questions or concerns about your rights in this research study, please contact the Clemson University Office of Research Compliance (ORC) at [irb@clemson.edu](mailto:irb@clemson.edu).

You may print a copy of this informational letter for your files.

### **Study 3 Debriefing:**

Additional Information:

Thank you for taking part in this study. You were told at the beginning of the study that the purpose of this study was to examine how well untrained raters are able to score job interviews. Now that you have completed the study, we want to let you know that the true purpose of this study was actually to examine how the use of facial cosmetics and interviewee performance influence employment interview ratings. We did not tell you the true purpose of this study because we did not want you to focus on the candidate's appearance more or less than you naturally would.

If you would like a copy of the results of the study once it is completed, you may contact Skye Gillispie at [sgillis@clemson.edu](mailto:sgillis@clemson.edu).

Thank you again for taking part in this study!

## **Appendix E**

### **Video URLs**

#### **Pilot Study 2 Video URLs:**

Cosmetics Condition 1: [vimeo.com/235277697](https://vimeo.com/235277697)

Cosmetics Condition 2: [vimeo.com/235277149](https://vimeo.com/235277149)

Cosmetics Condition 3: [vimeo.com/235277709](https://vimeo.com/235277709)

Cosmetics Condition 4: [vimeo.com/235277813](https://vimeo.com/235277813)

Cosmetics Condition 5: [vimeo.com/235277832](https://vimeo.com/235277832)

#### **Study 3 Video URLs:**

Cosmetics Condition 1 – Low Performance – Question 1: <https://youtu.be/JtFfJToBIFU>

Cosmetics Condition 1 – Low Performance – Question 2: <https://youtu.be/jALEskOjaYA>

Cosmetics Condition 1 – Low Performance – Question 3:

<https://youtu.be/zTmmjTyhrWU>

Cosmetics Condition 1 – Intermediate Performance – Question 1: <https://youtu.be/-WErddSZtYs>

Cosmetics Condition 1 – Intermediate Performance – Question 2:

<https://youtu.be/ZCwJalc9HjI>

Cosmetics Condition 1 – Intermediate Performance – Question 3:

<https://youtu.be/VKCWO5zSZfU>

Cosmetics Condition 1 – High Performance – Question 1:

<https://youtu.be/dAuagpTXBUo>

Cosmetics Condition 1 – High Performance – Question 2:

<https://youtu.be/hmt4nXAXac4>

Cosmetics Condition 1 – High Performance – Question 3:

<https://youtu.be/CKTsOo4gNf8>

Cosmetics Condition 2 – Low Performance – Question 1: <https://youtu.be/dM23LI0oLVg>

Cosmetics Condition 2 – Low Performance – Question 2: <https://youtu.be/3TyEpgiXOdE>

Cosmetics Condition 2 – Low Performance – Question 3: <https://youtu.be/ahwtlbpPPA>

Cosmetics Condition 2 – Intermediate Performance – Question 1: <https://youtu.be/ycl8x-WW04I>

Cosmetics Condition 2 – Intermediate Performance – Question 2:

<https://youtu.be/Wrfbe5YVBrY>

Cosmetics Condition 2 – Intermediate Performance – Question 3:

<https://youtu.be/dMhawzrUPdU>

Cosmetics Condition 2 – High Performance – Question 1: <https://youtu.be/stf7uFegbBY>

## Appendix E (continued)

Cosmetics Condition 2 – High Performance – Question 2: <https://youtu.be/Ejl9H1KT9Dg>  
Cosmetics Condition 2 – High Performance – Question 3: <https://youtu.be/i9vClh-RbD8>

Cosmetics Condition 3 – Low Performance – Question 1: <https://youtu.be/DbMYNnfCs4>  
Cosmetics Condition 3 – Low Performance – Question 2: <https://youtu.be/Kiktx6R6Nus>  
Cosmetics Condition 3 – Low Performance – Question 3: <https://youtu.be/Vh4EtRpSDwk>

Cosmetics Condition 3 – Intermediate Performance – Question 1:  
<https://youtu.be/ihHHyGvHOM4>  
Cosmetics Condition 3 – Intermediate Performance – Question 2:  
<https://youtu.be/52afdgOlQkI>  
Cosmetics Condition 3 – Intermediate Performance – Question 3:  
<https://youtu.be/WhUh2De-Qo8>

Cosmetics Condition 3 – High Performance – Question 1: <https://youtu.be/tjcRbAKclB4>  
Cosmetics Condition 3 – High Performance – Question 2: <https://youtu.be/CMqzE0qfATc>  
Cosmetics Condition 3 – High Performance – Question 3: <https://youtu.be/xZr248WvqNI>

## Tables

*Table 1. Pilot Study 2- Descriptive statistics for the dependent variable, “The candidate's makeup was applied well.”*

Cosmetics Condition	N	Mean	SD	SE
2	50	4.46	.65	.09
3	50	4.52	.81	.12
4	46	4.50	.75	.11
5	54	4.15	.94	.13
Total	251	4.35	.84	.05



*Table 2. Pilot Study 2- Descriptive statistics for the dependent variable, “It was obvious that the candidate was wearing makeup.”*

Cosmetics				
Condition	N	Mean	SD	SE
1	51	2.86	1.18	.17
2	50	3.20	1.05	.15
3	50	3.50	1.18	.17
4	46	3.98	.91	.13
5	54	4.28	1.02	.14
Total	251	3.57	1.17	.08

*Table 3. Pilot Study 2- Descriptive statistics for the dependent variable, “How much makeup would you say the candidate was wearing?”*

Cosmetics				
Condition	N	Mean	SD	SE
1	51	2.14	.87	.12
2	50	2.46	.65	.09
3	50	2.62	.73	.10
4	46	2.87	.72	.11
5	54	3.30	.92	.13
Total	251	2.68	.88	.06

*Table 4. Study 3- Descriptive statistics for all continuous variables*

---

	N	Mean	SD
Hiring Recommendation	452	4.21	.80
Interview Score	452	3.60	.95
Attractiveness	452	3.76	.81
Professional Appearance	452	4.63	.52
Competence	452	4.37	.82

---

*Table 5. Study 3- Correlations between continuous variables.*

	1.	2.	3.	4.
1. Hiring Recommendation	1			
2. Interview Score	.455**	1		
3. Attractiveness	.280**	.218**	1	
4. Professional Appearance	.408**	.114*	.136**	1
5. Competence	.556**	.360**	.234**	.342**

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

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

*Table 6. Study 3- Mean interview ratings by cosmetics condition.*

---

Cosmetics Condition	N	Mean	SD	SE
1	154	3.66	.90	.07
2	151	3.55	.98	.08
3	147	3.57	.96	.08
Total	452	3.60	.95	.04

---

*Table 7. Research Question 1 - Interview rating by performance and cosmetics conditions.*

Performance	Cosmetics	Mean	SD	N
1	1	3.01	.76	52
	2	2.62	.78	46
	3	2.66	.72	51
	Total	2.77	.77	149
2	1	3.50	.57	52
	2	3.56	.60	54
	3	3.67	.62	44
	Total	3.57	.59	150
3	1	4.52	.59	50
	2	4.38	.68	51
	3	4.38	.56	52
	Total	4.42	.61	153
Total	1	3.66	.90	154
	2	3.55	.98	151
	3	3.57	.96	147
	Total	3.60	.95	452

## Figures

*Figure 1.* Interview structure levels from Huffcutt and Arthur (1994).

		Interview Question Standardization			
		1	2	3	4
Response Scoring Standardization	1	Structure 1	Structure 2	Structure 2	Structure 3
	2	Structure 2	Structure 2	Structure 3	Structure 3
	3	Structure 2	Structure 3	Structure 3	Structure 4

Note: “Level 1 question standardization was no constraints; Level 2 was limited constraints, typically on the topical areas; Level 3 was precise specification of questions from which interviewers could choose or follow-up; Level 4 was asking the exact same questions with no choice or follow-up. Level 1 response scoring was a global assessment; Level 2 response scoring was assessment along multiple established criteria; Level 3 was evaluation of each individual response according to preestablished answers.” (p. 187)

Figure 2. Huffcutt et al. (2011) Model of interviewee performance as a mediating construct between candidate attributes and interviewer ratings.

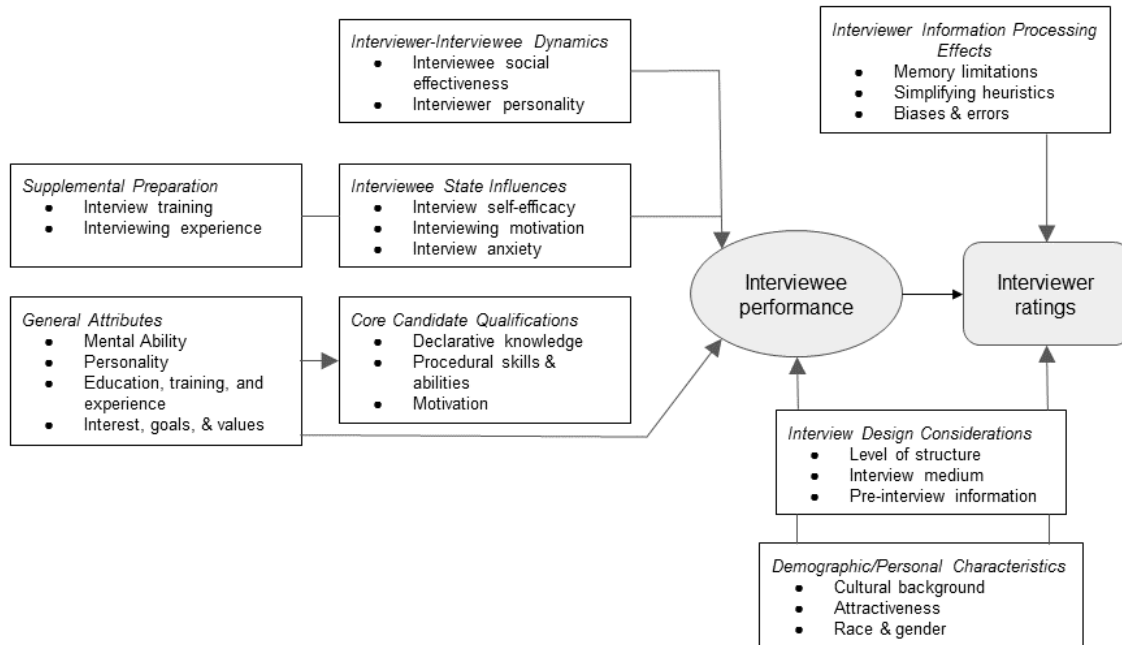




Figure 3. Hayes and Preacher (2014) mediation model with a multicategorical independent variable.

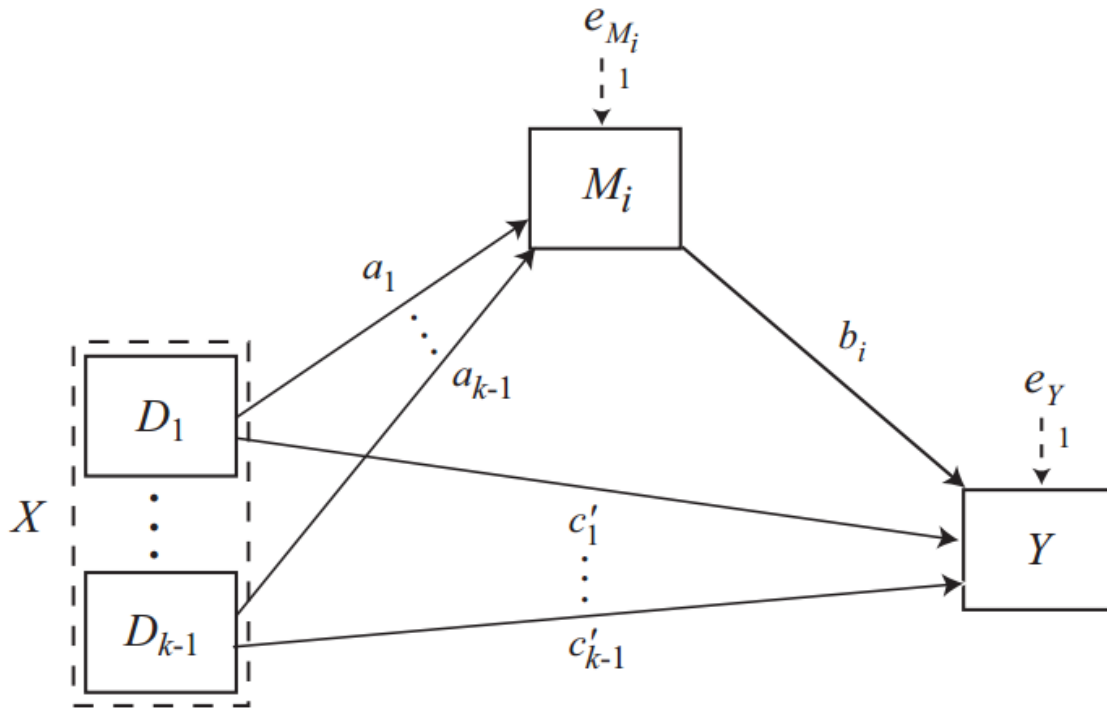
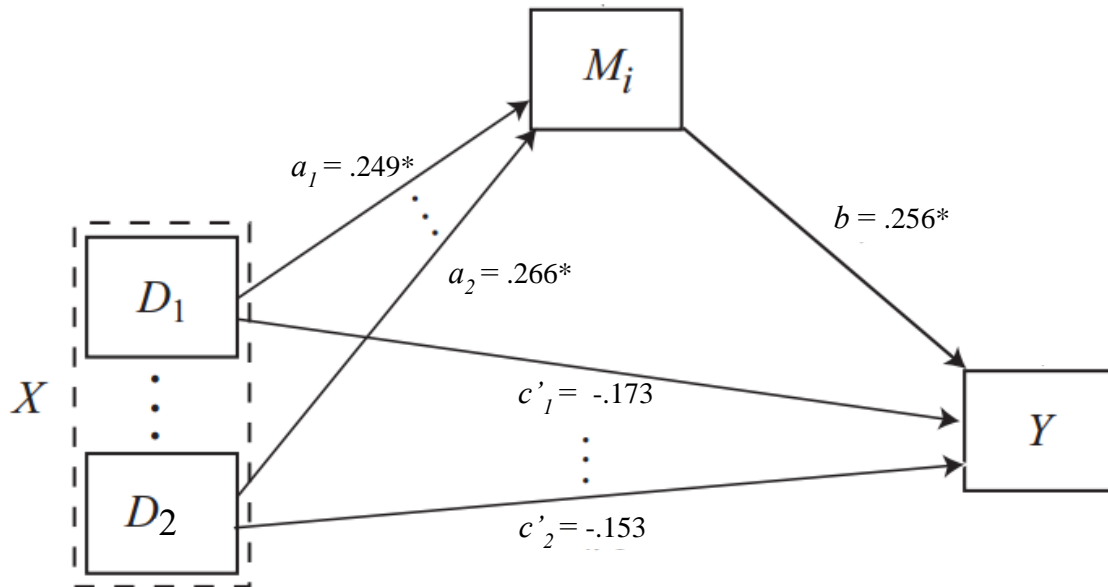


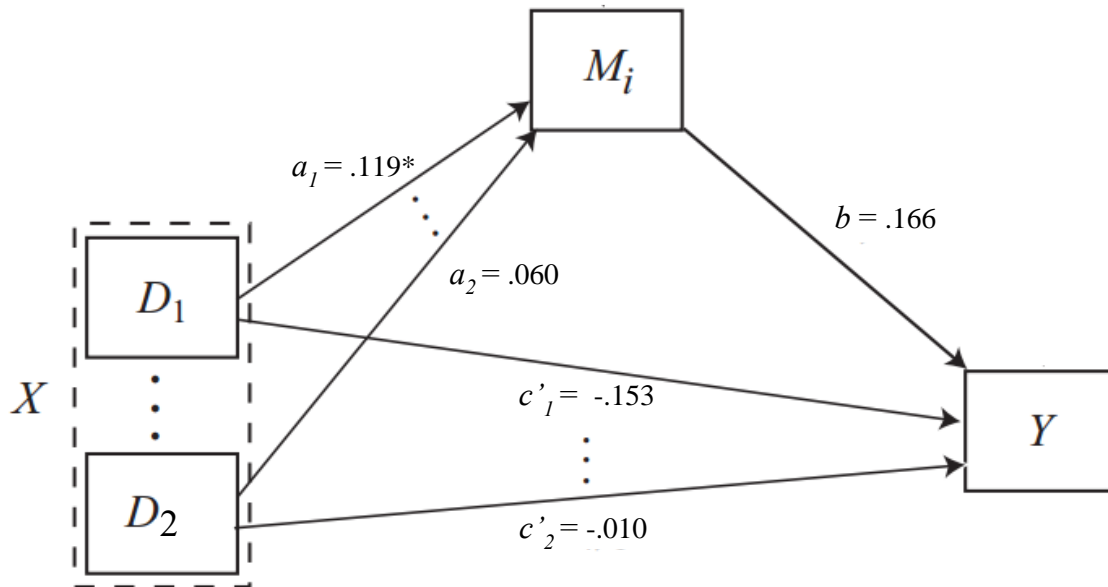
Figure 4. Hypothesis 1c mediation model.



Note: D1 is dummy coded for cosmetics condition 2. D2 is dummy coded for cosmetics condition 3. M is physical attractiveness, and Y is interview ratings.

\* indicates that the effect was significant at the  $p < .05$  level.

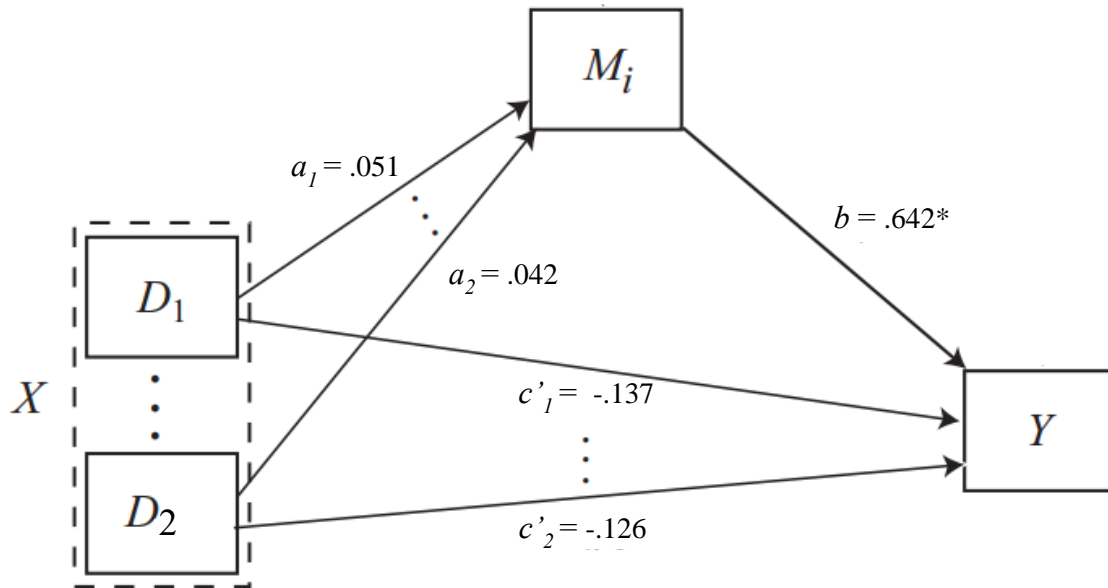
Figure 5. Hypothesis 2b mediation model.



Note: D1 is dummy coded for cosmetics condition 2. D2 is dummy coded for cosmetics condition 3. M is professional appearance, and Y is interview ratings.

\* indicates that the effect was significant at the  $p < .05$  level.

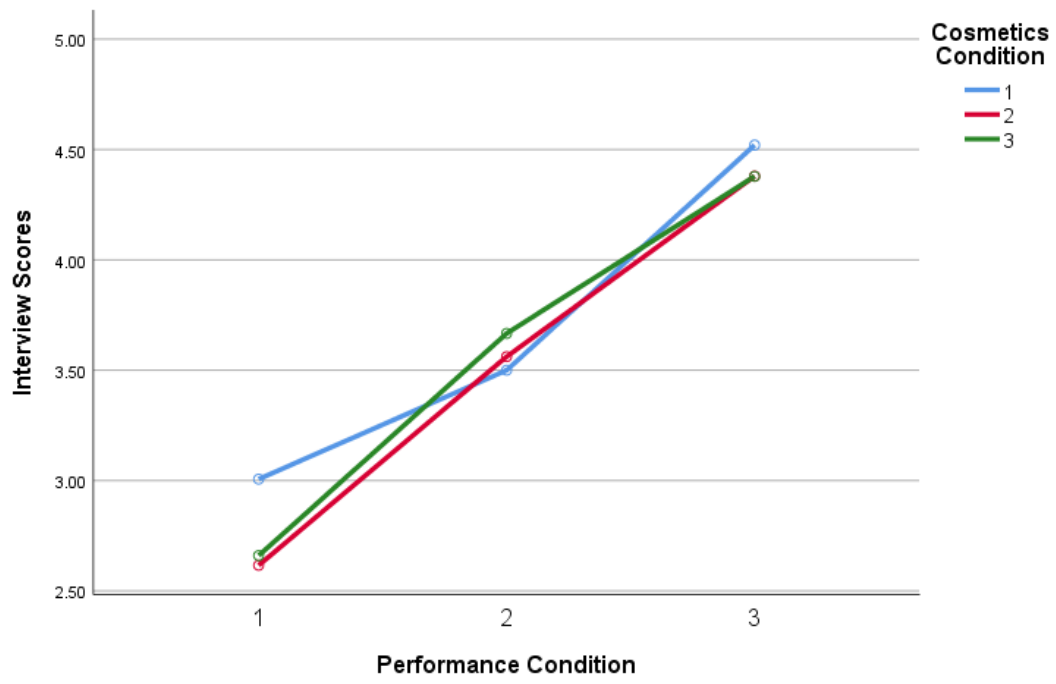
Figure 6. Hypothesis 3b mediation model.



Note: D1 is dummy coded for cosmetics condition 2. D2 is dummy coded for cosmetics condition 3. M is perceived competence, and Y is interview ratings.

\* indicates that the effect was significant at the  $p < .05$  level.

Figure 7. Interaction between performance and cosmetics on interview scores.



## References

- Arthur, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435-442. doi: 10.1037/0021-9010.93.2.435
- Barrick, M. R., Dustin, S. L., Giluk, T. L., Stewart, G. L., Shaffer, J. A., & Swider, B. W. (2012). Candidate characteristics driving initial impressions during rapport building: Implications for employment interview validity. *Journal of Occupational and Organizational Psychology, 85*(2), 330-352. doi:10.1111/j.2044-8325.2011.02036.x
- Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology, 94*(6), 1394-1411. doi:10.1037/a0016532
- Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial evaluations in the interview: Relationships with subsequent interviewer evaluations and employment offers. *Journal of Applied Psychology, 95*(6), 1163-1172. doi:10.1037/a0019918
- Bragger, J. D., Kutcher, E., Morgan, J., & Firth, P. (2002). The effects of the structured interview on reducing biases against pregnant job applicants. *Sex Roles, 46*(7-8), 215-226. doi:10.1023/A:1019967231059
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science, 6*(1), 3-5. doi:10.1177/1745691610393980
- Campion, M. A., Campion, J. E., & Hudson, J. J. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79*(6), 998-1002. doi:10.1037/0021-9010.79.6.998
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*(3), 655-702. doi:10.1111/j.1744-6570.1997.tb00709.x
- Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41*(1), 25-42.

- Cardellino, C. (2013, December 13). The perfect hair and makeup to nail your dream job interview. Retrieved from <http://www.cosmopolitan.com/style-beauty/beauty/how-to/a17046/cosmo-beauty-advice-column-2/>
- Cash, T. F., Dawson, K., Davis, P., Bowen, M., & Galumbeck, C. (1989). Effects of cosmetics use on the physical attractiveness and body image of American college women. *The Journal of Social Psychology, 129*(3), 349-355. doi:10.1080/00224545.1989.9712051
- Chapman, D. S., & Zweig, D. I. (2005). Developing A nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology, 58*(3), 673-702. doi:10.1111/j.1744-6570.2005.00516.x
- Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*(5), 565-579. doi:10.1037/0021-9010.80.5.565
- Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business And Psychology, 13*(4), 485-506. doi:10.1023/A:1022914803347
- Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*(2), 325-351. doi:10.1111/j.1744-6570.2000.tb00204.x
- Cox, C. L., & Glick, W. H. (1986). Resume evaluations and cosmetics use: When more is not better. *Sex Roles, 14*(1-2), 51-58. doi:10.1007/BF00287847
- Crissy, W. E., & Regan, J. J. (1951). Halo in the employment interview. *Journal of Applied Psychology, 35*(5), 338-341. doi:10.1037/h0060581
- Cunningham, J., & Macan, T. (2007). Effects of applicant pregnancy on hiring decisions and interview ratings. *Sex Roles, 57*(7-8), 497-508. doi:10.1007/s11199-007-9279-0
- Dellinger, K., & Williams, C. L. (1997). Makeup at work: Negotiating appearance rules in the workplace. *Gender & Society, 11*(2), 151-177. doi:10.1177/089124397011002002
- Del Russo, M. (2015, September 15). Get hired on the spot with these interview-appropriate beauty tips. Retrieved from <https://www.popsugar.com/beauty/Makeup-Tips-Job-Interview-1083269>

- Deprez-Sims, A., & Morris, S. B. (2013). The effect of non-native accents on the evaluation of applicants during an employment interview: The development of a path model. *International Journal of Selection and Assessment*, 21(4), 355-367. doi:10.1111/ijsa.12045
- Derous, E., Buijsrogge, A., Roulin, N., & Duyck, W. (2016). Why your stigma isn't hired: A dual-process framework of interview bias. *Human Resource Management Review*, 26(2), 90-111. doi:10.1016/j.hrmr.2015.09.006
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3), 285-290. doi:10.1037/h0033731
- Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110(1), 109-128. doi:10.1037/0033-2909.110.1.109
- Eder, R. W., & Harris, M. M. (1999). *The employment interview handbook*. Thousand Oaks, California: SAGE Publications.
- Eldridge, L. (2015). *Face paint The story of makeup*. New York: Abrams Image.
- Ellis, A. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type?. *Journal of Applied Psychology*, 87(6), 1200-1208. doi:10.1037/0021-9010.87.6.1200
- Etcoff, N. L., Stock, S., Haley, L.E., Vickery, S.A., & House, D. M. (2011) Cosmetics as a feature of the extended human phenotype: Modulation of the perception of biologically important facial signals. *PLoS ONE* 6(10), 1-9. doi:10.1371/journal.pone.0025656
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, 111(2), 304-341. doi:10.1037/0033-2909.111.2.304
- Fink, B., Grammer, K., & Thornhill, R. (2001). Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology*, 115(1), 92-99. doi:10.1037/0735-7036.115.1.92



- Fletcher, T. D., & Major, D. A. (2006). The effects of communication modality on performance and self-ratings of teamwork components. *Journal of Computer-Mediated Communication, 11*(2), 557-576. doi:10.1111/j.1083-6101.2006.00027.x
- Frieder, R. K., Van Iddekinge, C. H., & Raymark, P. H. (2016). How quickly do interviewers reach decisions? An examination of interviewers' decision-making time across applicants. *Journal of Occupational and Organizational Psychology, 89*, 223-248.
- Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. *Journal of Educational Statistics, 1*(2), 113-125. doi:10.2307/1164979
- Guéguen, N. (2008). Brief report: The effects of women's cosmetics on men's approach: An evaluation in a bar. *North American Journal of Psychology, 10*(1), 221-228.
- Gold, G. (2011, June 2). Makeup tips for facial symmetry. Retrieved from <http://www.youbeauty.com/beauty/makeup-tips-for-facial-symmetry/>
- Gonin, L. –C. M. W. (1945). 'Report on the Liberation of Belsen', Imperial War Museum, Department of Documents, File 85/38/1.
- Hartwell, C. J., & Campion, M. A. (2016). Getting on the same page: The effect of normative feedback interventions on structured interview ratings. *Journal of Applied Psychology, 101*(6), 757-778. doi:10.1037/apl0000099
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*(3), 639-683. doi:10.1111/j.1744-6570.2004.00003.x
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67*(3), 451-470. doi:10.1111/bmsp.12028
- Highhouse, S., Doverspike, D., & Guion, R. M. (2016). *Essentials of personnel assessment and selection., 2nd ed.* New York, NY, US: Routledge/Taylor & Francis Group.
- Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology, 56*(2), 431-462. doi:10.1111/j.1744-6570.2003.tb00157.x

- Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*(2), 184-190. doi:10.1037/0021-9010.79.2.184
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*(5), 897-913. doi:10.1037/0021-9010.86.5.897
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment, 21*, 264-276.
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment, 22*(3), 297-309. doi:10.1111/ijsa.12078
- Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*(2), 179-189. doi:10.1037/0021-9010.83.2.179
- Huffcutt, A. I., Roth, P. L., Conway, J. M., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*(5), 897-913. doi:10.1037//0021-9010.86.5.897
- Huffcutt, A. I., Van Iddekinge, C. H., & Roth, P. L. (2011). Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. *Human Resource Management Review, 21*(4), 353-367.
- Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology, 54*(3), 619-644. doi:10.1111/j.1744-6570.2001.tb00225.x
- Huguet, P., Croizet, J., & Richetin, J. (2004). Is 'What has been cared for' necessarily good? Further evidence for the negative impact of cosmetics use on impression formation. *Journal of Applied Social Psychology, 34*(8), 1752-1771. doi:10.1111/j.1559-1816.2004.tb02796.x
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72-98. doi:10.1037/0033-2909.96.1.72

- Ingold, P. V., Kleinmann, M., König, C. J., Melchers, K. G., & Van Iddekinge, C. H. (2015). Why do situational interviews predict job performance? The role of interviewees' Ability to Identify Criteria. *Journal of Business and Psychology, 30*(2), 387-398. doi:10.1007/s10869-014-9368-3
- Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. *Journal of Applied Psychology, 64*(5), 502-508. doi:10.1037/0021-9010.64.5.502
- Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly, 58*(2), 108-122. doi:10.2307/2787149
- Jones, A. L., Russell, R., & Ward, R. (2015). Cosmetics alter biologically-based factors of beauty: Evidence from facial contrast. *Evolutionary Psychology, 13*(1), 210-229. doi:10.1177/147470491501300113
- Khazan, O. (2015, August 5). The makeup tax. Retrieved from <https://www.theatlantic.com/business/archive/2015/08/the-makeup-tax/400478/>
- Kinicki, A. J., & Lockwood, C. A. (1985). The interview process: An examination of factors recruiters use in evaluating job applicants. *Journal of Vocational Behavior, 26*(2), 117-125. doi:10.1016/0001-8791(85)90012-0
- Klehe, U., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance, 21*(2), 107-137. doi:10.1080/08959280801917636
- Klehe, U., & Latham, G. (2006). What would you do--really or ideally? Constructs underlying the behavior description interview and the situational interview in predicting typical versus maximum performance. *Human Performance, 19*, 357-382. doi:10.1207/s15327043hup1904\_3
- Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology, 100*(1), 128-161. doi:10.1037/a0036734
- Landy, F. J. (2008). Stereotyping, implicit association theory, and personnel decisions: I guess we will just have to agree to disagree. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*(4), 444-453. doi:10.1111/j.1754-9434.2008.00085.x

- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390-423. doi:10.1037/0033-2909.126.3.390
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241-293. doi:10.1111/peps.12052
- Lievens, F., Highhouse, S., & DeCorte, W. (2005). The importance of traits and abilities in supervisors' hirability decisions as a function of method of assessment. *Journal of Occupational and Organizational Psychology*, *78*(3), 453-470. doi:10.1348/096317905X26093
- Lin, T., Dobbins, G. H., & Farh, J. (1992). A field study of race and age similarity effects on interview ratings in conventional and situational interviews. *Journal of Applied Psychology*, *77*(3), 363-371. doi:10.1037/0021-9010.77.3.363
- Mack, D., & Rainey, D. (1990). Female applicants' grooming and personnel selection. *Journal of Social Behavior & Personality*, *5*(5), 399-407.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, *79*(4), 599-616. doi:10.1037/0021-9010.79.4.599
- Melchers, K. G., Klehe, U., Richter, G. M., Kleinmann, M., König, C. J., & Lievens, F. (2009). 'I know what you want to know': The impact of interviewees' ability to identify criteria on interview performance and construct-related validity. *Human Performance*, *22*(4), 355-374. doi:10.1080/08959280903120295
- Mileva, V. R., Jones, A. L., Russell, R., & Little, A. C. (2016). Sex differences in the perceived dominance and prestige of women with and without cosmetics. *Perception*, *45*(10), 1166-1183. doi:10.1177/0301006616652053
- Miller, L. C., & Cox, C. L. (1982). For appearances' sake: Public self-consciousness and makeup use. *Personality and Social Psychology Bulletin*, *8*(4), 748-751.
- Morgan, W. B., Walker, S. S., Hebl, M. R., & King, E. B. (2013). A field experiment: Reducing interpersonal discrimination toward pregnant job applicants. *Journal of Applied Psychology*, *98*(5), 799-809. doi:10.1037/a0034040
- Moscoso, S. (2000). A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment*, *8*, 237-247. doi:10.1111/1468-2389.00153

- Motowidlo, S. J., & Burnett, J. R. (1995). Aural and visual sources of validity in structured employment interviews. *Organizational Behavior and Human Decision Processes*, 61(3), 239-249. doi:10.1006/obhd.1995.1019
- Mulhern, R., Fieldman, G., Hussey, T., Leveque, J. L., & Pineau, P. (2003). Do cosmetics enhance female Caucasian facial attractiveness? *International Journal of Cosmetic Science*, 25, 199-205. DOI: 10.1046/j.1467-2494.2003.00188.x
- Nash, R., Fieldman, G., Hussey, T., L  v  que, J., & Pineau, P. (2006). Cosmetics: They Influence More Than Caucasian Female Facial Attractiveness. *Journal of Applied Social Psychology*, 36(2), 493-504. doi:10.1111/j.0021-9029.2006.00016.x
- O'Brien, J., & Rothstein, M. G. (2011). Leniency: Hidden threat to large-scale, interview-based selection systems. *Military Psychology*, 23(6), 601-615. doi:10.1080/08995605.2011.616791
- Oostrom, J. K., Melchers, K. G., Ingold, P. V., & Kleinmann, M. (2016). Why do situational interviews predict performance? Is it saying how you would behave or knowing how you should behave?. *Journal of Business and Psychology*, 31(2), 279-291. doi:10.1007/s10869-015-9410-0
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.
- Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior*, 20(5), 295-307. doi:10.1016/S1090-5138(99)00014-8
- Piess, K. (1998). *Hope in a jar: The making of America's beauty culture*. New York: Henry Holt Company.
- Pingitore, R., Dugoni, B. L., Tindale, R. S., & Spring, B. (1994). Bias against overweight job applicants in a simulated employment interview. *Journal of Applied Psychology*, 79(6), 909-917. doi:10.1037/0021-9010.79.6.909
- Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, 55(1), 1-81. doi:10.1111/j.1744-6570.2002.tb00103.x
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11(1), 37-43. doi:10.1016/j.tics.2006.11.001

- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28(3), 369-381. doi:10.1177/0146167202286008
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69(4), 581-588. doi:10.1037/0021-9010.69.4.581
- Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel Psychology*, 48(2), 289-308.
- Purkiss, S. S., Perrewé, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G. R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes*, 101(2), 152-167. doi:10.1016/j.obhdp.2006.06.005
- Ramien, M. L., Ondrejchak, S., Gendron, R., Hatami, A., McCuaig, C., Powell, J., & Marcoux, A. (2014). Quality of life in pediatric patients before and after cosmetic camouflage of visible skin conditions. *American Academy of Dermatology*, 71, 935-940. Doi: <http://dx.doi.org/10.1016/j.jaad.2014.07.029>
- Raza, S. M., & Carpenter, B. N. (1987). A model of hiring decisions in real employment interviews. *Journal of Applied Psychology*, 72(4), 596-603. doi:10.1037/0021-9010.72.4.596
- Robertson, J., Fieldman, G., & Hussey, T. (2008). 'Who wears cosmetics?' Individual differences and their relationship with cosmetic usage. *Individual Differences Research*, 6(1), 38-56.
- Roth, P. L., & Huffcutt, A. I. (2013). A meta-analysis of interviews and cognitive ability: Back to the future?. *Journal of Personnel Psychology*, 12, 157-169. doi:10.1027/1866-5888/a000091
- Ruetzler, T., Taylor, J., Reynolds, D., Baker, W., & Killen, C. (2012). What is professional attire today? A conjoint analysis of personal presentation attributes. *International Journal of Hospitality Management*, 31(3), 937-943. doi:10.1016/j.ijhm.2011.11.001
- Russell, R. (2010). Why cosmetics work. In R. B. Adams, A. Ambady, K. Nakayam, & S. Shimojo (Eds.), *The science of social vision* (pp. 186-204). Oxford: Oxford University Press.

- Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology, 11*, 299-324. doi:10.1080/13594320244000184
- Salihu, H. M., Myers, J., & August, E. M. (2012). Pregnancy in the workplace. *Occupational Medicine, 62*(2), 88-97. doi:10.1093/occmed/kqr198
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274. doi:10.1037/0033-2909.124.2.262
- Schmidt, F. L., & Zimmerman, R. D. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology, 89*(3), 553-561. doi:10.1037/0021-9010.89.3.553
- Shahani, C., Dipboye, R. L., & Gehrlein, T. M. (1991). The incremental contribution of an interview to college admissions. *Educational and Psychological Measurement, 51*(4), 1049-1061. doi:10.1177/001316449105100425
- Scott, W. D. (1915). Scientific selection of salesman. *Advertising and Selling Magazine, 5*, 5-6.
- Squier, C. (2016, April 27). The male make-up industry is growing – Here’s why that’s a good thing. Retrieved from <http://www.thedebrief.co.uk/style/hair-and-beauty/male-make-up-20160463196>
- Tagai, K., Ohtaka, H., & Nittono, H. (2016). Faces with light makeup are better recognized than faces with heavy makeup. *Frontiers in Psychology, 7*, 1-8. doi:10.3389/fpsyg.2016.00226
- Topor, D. J., Colarelli, S. M., & Han, K. (2007). Influences of traits and assessment methods on human resource practitioners' evaluations of job applicants. *Journal of Business and Psychology, 21*(3), 361-376. doi:10.1007/s10869-006-9033-6
- Tsai, W., Chen, C., & Chiu, S. (2005). Exploring Boundaries of the Effects of Applicant Impression Management Tactics in Job Interviews. *Journal of Management, 31*(1), 108-125. doi:10.1177/0149206304271384
- Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., Jr., Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance, 17*, 71-93. doi:10.1207/S15327043HUP1701\_4

- Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*(3), 536-552. doi:10.1037/0021-9010.90.3.536
- Walters, L. C., Miller, M. R., & Ree, M. J. (1993). Structured interviews for pilot selection: No incremental validity. *The International Journal of Aviation Psychology, 3*(1), 25-38. doi:10.1207/s15327108ijap0301\_2
- Welch, B. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38*(3/4), 330-336. doi:10.2307/2332579
- Wexley, K. N., Sanders, R. E., & Yukel, G. A. (1973). Training interviewers to eliminate contrast effects in employment interviews. *Journal of Applied Psychology, 57*(3), 233-236. doi:10.1037/h0034714
- Whipp, L. (2017, February 4). Made-up men reflect changing \$50bn male grooming industry. Retrieved from <https://www.ft.com/content/825e520c-c798-11e6-8f29-9445cac8966f>
- Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 61*(4), 275-290. doi:10.1111/j.2044-8325.1988.tb00467.x
- Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., & Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology, 82*(6), 900-912. doi:10.1037/0021-9010.82.6.900
- Wong, J. S., & Penner, A. M. (2016). Gender and the returns to attractiveness. *Research in Social Stratification and Mobility, 44*, 113-123.
- Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of Occupational Psychology, 62*(3), 191-199.