Data Analytics and Skeptical Actions: The Countervailing Effects of False Positives and Consistent Rewards for Skepticism

By: Dereck Barr-Pulliam, Joseph Brazel, Jennifer McCallen and Kimberly Walker

ABSTRACT: We investigate if varying rates of false positives impact auditor skepticism toward red flags identified by data analytic tools. We also examine the extent to which consistent rewards for skepticism can improve the application of skepticism on audits employing data analytics. Using an experiment with practicing auditors we observe that when false positive rates are higher, skepticism levels are low. We also find that consistent rewards for skepticism significantly improve the skepticism of our auditors. However, the positive effect of rewards is limited in that we do not see improvements in skepticism when the false positive rate is higher and additional investigation is less likely to identify a misstatement. Our findings highlight the importance of calibrating analytic tools to reduce false positives and the need for a culture change where appropriate skepticism is consistently rewarded in order for audit firms to effectively use analytic tools to enhance audit quality.

Keywords: audit; data analytics; false positives; professional skepticism; rewards

Data Availability: Contact the authors

I. INTRODUCTION

The emergence of new data analytic techniques in the audit environment allows auditors to gain deeper insights into their clients' data, but simultaneously creates unique challenges for auditors when exercising skepticism (Austin, Carpenter, Christ, and Neilson [2019], Arens, Elder, Beasley, and Hogan [2020]). Data analytics facilitate auditors' understanding of rich data sets by identifying patterns and relationships and helping auditors to detect anomalies within the data (AICPA [2017]). Because data analytic approaches enable auditors to examine full populations (versus sampling) and incorporate more diverse data and visualizations into their testing, auditors are often faced with a larger number of anomalies or relationships that should be investigated and further evaluated (No, Lee, Huang, and Li [2019]). This creates a unique dilemma for auditors as they determine not only how to best utilize information from data analytic tools, but also how to manage the investigation of red flags identified by these tools.

One concern over the use of data analytics is the likelihood of *false positives*, or the extent that these tools identify transactions or relationships as potential anomalies that, after further investigation, are determined to be reasonable, explained variations in the data (AICPA [2017], Baadder and Krcmar [2018], Johnson and Wiley [2019], Austin et al. [2019], Brazel, Leiby, and Schaefer [2020], Arens et al. [2020]).¹ The frequency of false positives increases proportionately with the size and complexity of the data analyzed (Krahel and Titera [2015], Kogan, Mayhew, and Vasarhelyi [2019]). Despite concerns over false positive rates, few studies have addressed the problems auditors face when processing the outliers identified by analytic tools (No et al. [2019]).²

¹ For example, auditing textbooks are, for the first time, addressing the likelihood of false positives as a significant issue when using analytic tools during risk assessment and substantive testing (e.g., Johnson and Wiley [2019]; Arens et al. [2020]). Likewise, our discussions with practitioners confirm that false positive rates have become a significant issue that is considered when employing audit data analytics in practice (vs. when employing more traditional audit procedures).

 $^{^{2}}$ We use the terms evidence inconsistency, anomaly, red flag, outlier, and exception interchangeably. Each describes an instance where the auditor identifies a potential issue and decides whether to investigate it further.

Professional skepticism continues to be a focus of regulators, standard setters, practitioners, and researchers as it is essential to audit quality (e.g., Nelson [2009], PCAOB [2012], IAASB [2015] KPMG [2016], Nolder and Kadous [2018]). "Costly skepticism" occurs when auditors appropriately apply skepticism that generates additional costs (e.g., budget overages, strained client relations), but their efforts do not ultimately detect a misstatement (Brazel et al. [2020]). Audit supervisors typically do not reward costly skepticism and the anticipation that costly skepticism will not be rewarded can inhibit auditor skepticism when red flags or exceptions are encountered ([Brazel, Jackson, Schaefer, and Stewart [2016], Brown and Millar [2020]).³

Given the shift in the audit environment toward incorporating more data analytic approaches, it is important to identify and understand conditions in which the presence of data analytics influences how auditors apply professional skepticism. While the idea of costly skepticism is not new or specific to data analytics, the additional insights provided by analytic tools could dramatically increase the number of exceptions that ultimately become false positives. Auditors are therefore faced with the choice of potentially incurring costly skepticism to investigate red flags identified by analytic tools, or underutilizing data analytics and potentially failing to identify misstatements. We investigate how varying rates of false positives produced by analytic tools influence auditor skepticism and whether, and how, prior rewards for skepticism improve the application of skepticism on audits employing data analytics.

The false positive rate of data analytic tools is likely to influence the auditor's response to red flags identified by those tools. Exceptions generated by information technology are evaluated based on two message factors: perceived message severity and message ambiguity (Davis and

³ Skepticism that ultimately identifies a misstatement is also likely to generate costs, but such skepticism is framed by superiors as a *normal* cost of the audit that benefits the engagement team. As such, prior research observes that evaluators consistently reward skepticism that *does* identify a misstatement (Brazel et al. 2016; Brown and Millar 2020).

Tuttle [2013]). Experiencing higher false positive rates in the past is likely to decrease message severity and increase message ambiguity when encountering new exceptions, as dismissing the exception is less risky and there is more uncertainty about whether the exception will ultimately lead to an identified misstatement. Both of these perceptions should reduce the likelihood of investigation. Despite the necessity for auditors to be skeptical about red flags identified by analytic tools, higher false positive rates increase the risk that the auditor will ultimately engage in costly skepticism that will not be rewarded by their supervisor (Brazel et al. [2016]). As such, experiencing higher false positive rates in the past could lead to lower skepticism amongst auditors in subsequent tasks.

As noted previously, auditors are typically not rewarded for exercising costly skepticism or appropriately applying skepticism that incurs costs, but those efforts fail to ultimately identify a misstatement (Brazel et al. [2016], Brazel, Gimbar, Maskymov, and Schaefer [2019], Brown and Millar [2020]). However, more recent research suggests that if auditors are *consistently rewarded* for costly skepticism by their supervisors, they are more likely to act skeptically when presented with a red flag identified via traditional trend analyses (Brazel, et al. [2020]). Still, it is unknown how the presence of consistent rewards influences auditor skepticism when the auditor uses data analytics and the potential for false positives becomes a larger issue. Given the varied levels of false positives associated with analytic tools, and thus higher uncertainty about costly skepticism, the presence of rewards aimed at motivating skeptical action may not have the same impact.

In this study, we follow the guidance of Kogan et al. [2019] who describe how researchers in the field of audit data analytics should follow a "design science methodology" of identifying a problem, designing a novel solution, and validating if the solution is indeed successful. Specifically, we examine whether higher false positive rates cause auditors to be less responsive to red flags and whether a culture change that consistently rewards costly skepticism can improve auditors' skeptical actions. We also investigate if the presence of higher false positive rates ultimately diminishes the positive effects of rewards.

We conducted an experiment with practicing auditors who performed a year-end substantive analytical procedure for a hypothetical audit engagement. We manipulated two independent variables: (1) rewards for costly skepticism (consistently rewarding costly skepticism versus consistently *not* rewarding costly skepticism) and (2) the historical rate of false positives for analytic tools used on the engagement (lower versus higher). At the beginning of the task, auditors were informed that they had engaged in costly skepticism during interim testing and were either rewarded by receiving an "Exceeded Expectations" evaluation (consistently rewarded) or a "Met Expectations" evaluation and thus were not rewarded (consistently not rewarded). Further, all participants were informed that the evaluation was *consistent* with the evaluations they had received in the past from their manager when they had engaged in costly skepticism.

The case then asked the auditors to perform a year-end analytical procedure for sales based on a rich set of financial data and non-financial measures (NFMs, such as number of employees, patents, and products). We held sales growth constant, and the associated financial data (e.g., industry data, prior year trends) was consistent with the level of sales growth. However, NFM growth for the client was sharply negative in all cases, creating a red flag consistent with levels exhibited by companies that have committed financial statement fraud (Brazel, Jones, and Zimbelman [2009]). Participants received visualizations from a data analytic tool that analyzed the financial and NFM data from the last five years to assist in the development of their expectation

⁴ See Brazel et al. [2016] for a discussion of how supervisors consistently offer a "Met Expectations" evaluation when their subordinates engage in costly skepticism and how a "Met Expectations" evaluation is not perceived as a reward by subordinate auditors.

for the sales balance. The visualizations illustrated the NFM red flag for all participants, as well as the consistency found in the financial data.

While audit data analytics take many forms (e.g., visualizations, 100% sampling of transactions, regression analyses), we chose visualizations because they are employed by most audit firms, are relevant to our substantive analytical procedure task (Anderson, Hobson, and Peecher [2020]), and are featured prominently in the AICPA's *Guide to Audit Data Analytics* (AICPA [2017]).⁵ Further, practitioner interest in research examining the link between visualizations and skepticism is summarized by the following quote from a recent Center for Audit Quality listing of Topics of Interest: "Suggested areas for research include the potential role and uses of data visualization tools; how the integration of these various tools can enhance critical thinking and professional skepticism..."⁶

Prior to viewing the visualizations, the participants were informed that data analytic tools employed in the past on the engagement had a historical false positive rate of either 25% (lower false positive rate) or 75% (higher false positive rate).⁷ After reviewing all the provided information, participants completed the analytic procedure. Our primary measure of skepticism was whether or not the auditors concluded from the procedure that the sales account warranted additional investigation.

While false positives are typically associated with over-auditing or audit inefficiencies (e.g., Austin et al. [2019]), we observe *low* levels of professional skepticism amongst auditors when false positive rates are higher. Despite the presence and visualization of the fraud red flag,

⁵ For example, Anderson et al. [2020] employ histograms of accounts receivable partitioned by account magnitude and invoice age to simulate a data visualization, while the AICPA's *Guide to Audit Data Analytics* (AICPA [2017]) provides graphs of accounts receivable broken down by year and currency.

⁶ https://www.thecaq.org/rab-request-for-proposals-topics-of-interest-in-2020.

⁷ In the Method section we describe how, given the low base rate of misstated accounts (e.g., Durney, Elder, and Glover [2014]), both our lower and higher false positive conditions reflect settings that auditors would expect to encounter when using well calibrated analytic tools in the field.

only a third of participants indicated additional investigation was needed when the false positive rate was higher. We also find that consistent rewards for costly skepticism enhance auditor skepticism. Specifically, over half of the auditors who were rewarded concluded that the sales account required additional testing, whereas less than a third of the auditors who were *not* rewarded reached the same conclusion. However, we identify a boundary condition in that higher false positive rates significantly reduce the positive effect of rewards.

In additional analyses, we examine both the process by which false positives affect auditor skeptical actions and the link between skeptical judgments and skeptical actions proposed by Nelson [2009]. We find that the effect of the false positive rate on skepticism is mediated by message severity. When false positive rates rise, auditors are more comfortable dismissing the red flag identified by the analytic tool and are less apt to rely on the analytic and incorporate the inconsistent evidence into their testing. In turn, auditor skepticism decreases. We also find that, despite false positive rates and rewards impacting our auditors' skeptical actions, we do not observe any variation in their skeptical judgments. Auditors in all experimental conditions typically recognized the red flag, yet unless *both* false positives were lower and rewards were present, our auditors' skeptical judgments did not convert to skeptical actions (i.e., investigation of the red flag).

Further, even though auditors in all experimental conditions *recognized* the NFM red flag, we observe that when false positives are higher, auditors choose to *not document* the NFMs in their testing. This could perhaps be a strategic decision in case the audit is inspected or brought into litigation (i.e., the auditor could claim ignorance of the red flag and thus justify not investigating). However, not documenting the declining NFMs also reduces the likelihood that a superior or reviewer ever becomes aware of the fraud red flag and compromises the review process

as a quality control mechanism. Last, we demonstrate that auditor risk-aversion plays a role in how they respond to rewards and false positive levels.

Overall, our findings suggest that false positives associated with data analytics can influence the extent to which auditors exhibit professional skepticism. Auditing standards require auditors to exercise skepticism throughout the audit process [PCAOB 2019a], yet we find that auditors are less likely to act skeptically when data analytic techniques produce higher false positive rates. One deterrent to exhibiting skeptical behaviors is the lack of consistent rewards associated with costly skepticism (Brazel et al. [2016], Brazel et al. [2019]). Our results indicate that when auditors are consistently rewarded for exhibiting appropriate skepticism, regardless of the outcome, they are more likely to continue to act skeptically. Unfortunately, prior research has illustrated that such consistent rewards are rare (e.g., Brazel et al. [2016], Brazel et al. [2019]). Overall, we provide new evidence on the effects of data analytics in the audit environment, highlighting conditions in which auditors may underutilize these new tools, which could potentially threaten the effectiveness of audits. Our study highlights the importance of calibrating data analytic tools to reduce false positives and the need for a culture change where costly skepticism is consistently rewarded in order for audit firms to effectively use analytic tools to enhance audit quality.

II. BACKGROUND AND HYPOTHESES DEVELOPMENT

Data Analytics in Auditing

Audit data analytics "is the science and art of discovering and analyzing patterns, identifying anomalies, and extracting other useful information in data underlying or related to the subject matter of an audit through analysis, modeling, and visualization for the purpose of planning or performing the audit" AICPA [2017]. Austin et al. [2019] note that "audit firms are making substantial investments in technologies that will advance the use of data analytics in financial

statement audits, causing potentially critical changes to the traditional audit process." Data analytic tools have the ability to enhance the audit process. For example, data visualizations can help auditors better understand their clients' businesses by providing more meaningful insights into client data (EY [2014], KPMG 2014, PwC [2015], Deloitte [2016]). Other data analytic tools allow auditors to perform population testing which proportionately increases the sufficiency of audit evidence (e.g., Brown-Liburd, Issa, and Lombardi [2015]; Barr-Pulliam [2019]). Additionally, properly designed data analytic tools can increase the appropriateness of evidence because the tools improve the accuracy, timeliness, relevance, and breadth of evidence (Davidson, Desai, and Gerard [2013]). Data analytics can also increase the likelihood of identifying and testing audit areas associated with higher risk (e.g., KPMG [2014], PwC [2015]). Consistent with data analytics improving audit quality, a recent study finds that auditors' use of data analytic techniques can reduce the likelihood they are found contributorily negligent for an audit failure (Barr-Pulliam, Brown-Liburd, and Sanderson [2019]).

However, auditors often lack the technical expertise or knowledge to fully interact with emerging technologies (e.g., Walker, Liburd-Brown, and Lewis [2019]). As a result, audit firms and auditors often rely on the visual output and exception reports from these advanced tools (i.e., data visualizations) when performing audit tasks (e.g., Austin et al. [2019]). Generally, research provides evidence that support the benefits of data visualizations (e.g., Vessey [1991], Huang, Eades, and Hong [2009]). In an audit setting, Rose, Rose, Sanderson, and Thibodeau [2017]) find that when data visualizations supplement traditional audit procedures, auditors are better able to recognize patterns in data and incorporate that information into their judgments. However, auditors are still likely sensitive to the "newness" of data analytic techniques and: (1) may be unsure/inexperienced in how to properly assess and respond to output from data analytics and (2)

may struggle with the shift from traditional audit methods to data analytics (Brown-Liburd et al. [2015], Cao, Chychyla, and Stewart [2015], Yoon, Hoogduin, and Zhang [2015], Austin et al. [2019]). Indeed, a recent study observes that auditors are not apt to identify outliers as high risk when using data visualizations (histograms) during substantive analytical procedures (Anderson et al. [2020]).

False Positives and Costly Skepticism

Professional skepticism is essential to audit quality and enhancing auditor skepticism is of great concern to regulators, standard setters, practitioners, and researchers (e.g., Nelson [2009]; PCAOB [2012], IAASB [2015], KPMG [2016]). Despite its importance, little is known about the factors and conditions under which individual auditors have sufficient incentives to exercise skepticism (Hurtt, Brown-Liburd, Earley, and Krishnamoorthy [2013]).⁸ Skepticism that is *ex ante* appropriate based on the available evidence typically generates *ex post* costs (e.g., additional tests, audit report delays, budget overruns, or strained client relations). Further, because of the low base rate of misstatements (e.g., Durney, Elder, and Glover [2014]), applications of skepticism often do not identify a misstatement because additional investigations determine that there are reasonable explanations for any inconsistencies (Brazel et al. [2016]).

While auditors may leverage technology to enhance the sufficiency and appropriateness of their audit evidence, these data analytic techniques often provide results that are inconsistent with the client's reported financial data (Kogan et al. [2019]). Data analytic tools may identify evidence inconsistencies, exceptions, or red flags that suggest a higher risk of material misstatement, but investigations into such red flags are unlikely to result in an identified misstatement (Arens et al. [2020]). Still, auditors must initially determine the appropriate level of skepticism to apply to the

⁸ See also www.ifac.org/global-knowledge-gateway/audit-assurance/discussion/research-insights-auditor-professional.

investigation of these inconsistencies despite the likelihood that identified red flags may result in a large number of "false positives" (e.g., Schneider, Dai, Janvrin, Ajayi, and Raschke [2015]).

A false positive occurs when data analytic techniques identify exceptions, or red flags, that ultimately do not relate to an actual problem, risk, or misstatement (Johnson and Wiley [2019], Arens et al. [2020]). The false positive rate measures the number of false positives as a percentage of the total number of negative events or anomalies identified by the tool (Li, Chan, and Kogan [2016]). A higher (*lower*) rate of false positives indicates a higher (*lower*) percentage of instances where the data analytic tool (or set of tools) erroneously identifies a negative event. In their interviews of auditors in relation to data analytics, Austin et al. [2019] note that "auditors describe difficulty finding the right balance between spending time to make certain the anomalies are not indicative of material misstatement, and spending too much time on false positives and wasting valuable resources unnecessarily."⁹

Consider a situation where an auditor observes a potential red flag when using a data analytic tool and decides to exercise an appropriate level of skepticism by performing additional testing. The investigation of the red flag is consistent with the auditor appropriately applying skepticism; however, it requires added effort from both the auditor and the client, and such action does not necessarily ensure that a misstatement will be found. It is possible that the investigation leads to an acceptable explanation for the unusual pattern of facts observed, such that no audit adjustment is necessary. In short, the auditor has likely incurred the costs associated with

⁹ It is important to note that the false positive rate is not correlated with inherent risk. As stated previously, a false positive occurs when data analytic techniques identify exceptions, or red flags, that ultimately do not relate to an actual problem, risk, or misstatement (Johnson and Wiley [2019], Arens et al. [2020]). Inherent risk is the underlying risk a material misstatement that exists due to the nature and complexity of the account. The false positive rate is simply noise in the data or audit evidence and is unrelated to true misstatements. Therefore, the false positive rate is not indicative of, or a derivative of, more (or fewer) misstatements. In footnote 26 we illustrate how the auditor's perception of the risk of material misstatement does not mediate the effect of the false positive rate on auditor skepticism.

skepticism (e.g., went over budget, strained client relations), but has not experienced the "benefit" of identifying a misstatement.

We refer to appropriate skeptical behavior that generates costs, but does not identify a misstatement, as *costly skepticism* (Brazel et al. [2020]). Audit supervisors typically do not reward costly skepticism and the anticipation that skeptical actions will not be rewarded can inhibit auditor skepticism when red flags are encountered (Brazel et al. [2016], Brown and Millar [2020]). Researchers and practitioners have raised the possibility that a fear of false positives and, in turn, costly skepticism may cause auditors to underutilize data analytic tools (e.g., PwC [2014], Protiviti [2014], Wang and Cuthbertson [2015]).

From an auditing standpoint, red flags identified via data analytic techniques are no different than exceptions identified during traditional substantive testing. The auditor's response should not differ based on the format of identification and auditors are still required to exercise professional skepticism and address identified exceptions to determine whether there is an underlying misstatement in the financial statements (PCAOB [2019a]; Arens et al. [2020]). However, with the increasing availability of data and the ability of data analytic tools to analyze large datasets, auditors are more likely to encounter a larger number of exceptions and related false positives when using analytic tools (Krahel and Titera [2015]). Information technology research suggests that auditors' responses to exceptions identified via technology *will differ* from exceptions identified through traditional testing methods (Davis and Tuttle [2013]). Specifically, auditors' responses to red flags are likely to differ based on the false positive rate associated with analytic tools (Davis and Tuttle [2013]).

Davis and Tuttle [2013] develop a theoretical framework that examines the factors that influence users' responses to exceptions identified via information technology. Two exception

11

factors, perceived message severity and message ambiguity, play a role in determining the extent of the user's response to those exceptions. When message severity is low, dismissing the exception is the norm as users are less concerned with responding improperly (e.g., ignoring the red flag and not exercising skepticism). If message ambiguity is high, users must exert significant effort to simply determine if the message is credible, and thus, users are *less* likely to respond to exceptions (Davis and Tuttle [2013]).

Consistent with our setting, the false positive rate (higher vs. lower) may act as a cue related to the severity and ambiguity of identified exceptions. In an audit setting, when the false positive rate is higher, the perceived message severity is likely to be low. Because they have experienced a higher rate of false positives in the past, auditors will perceive a lower risk associated with dismissing red flags identified by analytic tools and will be less apt to rely on the analytic and take the red flag it highlights seriously. As a result, auditors will less apt to incorporate the inconsistent evidence into their testing and the likelihood of exercising professional skepticism will be lower.¹⁰ Further, a higher false positive rate may indicate higher ambiguity as to whether identified exceptions actually represent misstatements in the financial statements or unusual patterns of evidence that can be reasonably explained (Luippold and Kida [2012]). As the potential link between a red flag and a misstatement becomes more ambiguous and requires more effort to investigate, skeptical actions will diminish. After all, costly skepticism is typically not rewarded (Brazel et al. [2016]). As a result, we predict that when auditors are faced with higher false positive rates, they will be less likely to investigate a red flag identified by an analytic tool. Formally stated,

¹⁰ A common example of this notion is when a person receives an error message or threat warning from their computer and perhaps early on they respond to the message and investigate the issue (message severity is perceived to be high). However, if after several investigations they fail to see a tangible result, the individual will deem later alerts as less severe and will be more apt to ignore the message/not perceive a risk of failing to respond. In short, a modern day application of "The Boy Who Cried Wolf."

H1: When a red flag is identified by a data analytic tool, a higher rate of false positives will lead to less skeptical auditor actions.

The Effect of Rewards in the Presence of False Positives

Because audit supervisors respond negatively to budget overruns, costly skepticism can lead directly to poor performance on the primary objective measure that determines the auditor's evaluation (e.g., Agoglia, Hatfield, and Lambert [2015]). Brazel et al. [2016] and Brazel et al. [2019] provide evidence that audit supervisors *do not* consistently reward costly skepticism, even when they acknowledge that the skepticism is appropriate. Moreover, Brazel et al. [2016] illustrate that junior auditors *anticipate* that their supervisors will not reward costly skepticism, which should further limit skepticism in the field.

Given that evaluators are not apt to reward auditors when costly skepticism or false positives occur, Brazel et al. [2020] test the intuitive solution of implementing rewards for costly skepticism, as economic theory argues that rewarding a behavior should encourage that behavior (e.g., Prendergast [1999]). Importantly, Brazel et al. [2020] find that rewards for costly skepticism should be consistent, as a one-time reward actually *decreases* skepticism on subsequent tasks. Auditors view rewards for costly skepticism as uncommon and unlikely to be repeated in the future (Brazel et al. [2016]). As such, auditors consider a *one-time* reward as a better-than-expected outcome that leads auditors to approach subsequent tasks with a risk-averse gain frame (e.g., Tversky and Kahneman [1981]). Ultimately, auditors adopt a "quit while you're ahead" mentality, which heightens auditors' awareness to avoid the risks of further skeptical behavior and decreases skepticism toward more severe red flags, thus failing to exercise skepticism when doing so is most appropriate. Therefore, it is unlikely that any *one-time* reward would reduce concerns over incurring the costs of investigating red flags identified by analytic tools.

Brazel et al. [2020] do find evidence that experienced auditors who have worked under supervisors that *consistently* reward costly skepticism interpret rewards for costly skepticism as a credible signal, increasing their willingness to exercise skepticism on subsequent tasks as they believe the reward will be repeated. More importantly, auditors who have experienced a history of rewards for costly skepticism are more apt to apply skepticism when severe red flags are encountered. Still, given that such experiences were measured in the study (vs. manipulated in a controlled experiment as in this study), it is possible that the observed associations are the result of a confounded or correlated omitted variable.¹¹ It is also possible that the positive effects of consistent rewards do not convey to the use of data analytics, where the potential for false positives/risk of costly skepticism may play off auditors tendencies for risk-aversion (e.g., Lennox and Kauser [2017]).^{12,13} Thus, we explicitly examine the influence of consistent rewards on professional skepticism (versus the normal setting of consistently not rewarding costly skepticism). When utilizing a data analytic tool that identifies a red flag, we expect that auditors will be more apt to exhibit skeptical actions when they have been consistently rewarded for costly skepticism. Formally stated,

¹¹ For example, auditors that possess a higher level of audit expertise and more appropriately apply skepticism may also work for supervisors with greater expertise who are also more apt to reward costly skepticism. Indeed, auditor expertise could be an omitted variable in Brazel et al. [2020], as they observe that supervisors that are apt to reward costly skepticism are ranked higher than their peers.

¹² Brazel et al. [2020] examine the effects of one-time rewards in relation to a traditional analytical procedure (i.e., trend analysis). The study did not incorporate advanced data analytics (e.g., visualizations with Tableau) where the potential for false positives has become a larger issue (see footnote 1). Brazel et al. [2020] do perform an additional experiment with accounting *students* where consistent rewards for costly skepticism is manipulated and is found to have a marginal positive effect on skepticism; however, they note students may have different a priori reward expectations than experienced auditors. Our H2 explicitly tests the effect of consistent rewards on a sample of experienced auditors. Importantly, we also investigate a common condition where the positive rates (H3)).

¹³ Of course there is risk associated with *not* investigating a red flag identified by an analytic tool (i.e., failing to identify a material misstatement). Thus, risk-averse auditors may be apt to attend to such anomalies. However, Brazel et al. [2020] discuss how, in audit settings where the base rate of misstatements is low and costly skepticism is not rewarded, risk-averse auditors are likely to perceive that investigating a red flag as a risky endeavor (vs. the cautious route of not investigating). This would be especially be true in cases when the false positive rate is higher as we predict in H3.

H2: When a red flag is identified by a data analytic tool, consistently rewarding costly skepticism (vs. consistently not rewarding costly skepticism) will lead to more skeptical auditor actions.

Do Rewards Work when False Positive Rates are Higher?

Given the aforementioned negative effect that false positives may have on skeptical actions, it is possible that consistently rewarding skeptical behavior may not have the intended effect when data analytics are employed on an audit engagement. While Brazel et al. [2020] suggest that consistently rewarding costly skepticism will lead to increased skeptical actions when red flags are encountered, the use of data analytics presents a unique challenge to auditors. On the one hand, lower false positive rates may indicate higher levels of perceived message severity (i.e., a higher risk of failing to identify a misstatement by forgoing investigation) and lower levels of message ambiguity (i.e., a higher likelihood that identified exceptions are the result of an actual misstatement). As a result, when false positive rates are lower, concerns over costly skepticism are less likely and auditors should respond to consistent rewards by increasing skeptical behavior.

On the other hand, auditors tend to be risk-averse (e.g., Farmer [1993]; Davidson and Dalby [1993]; Lennox and Kausar [2017]). When false positive rates are higher, the risk that any acts of skepticism will *not* yield a misstatement becomes a larger concern for the auditor. Given a higher false positive rate, the cautious route may be to *not* exercise skepticism. Brazel et al. [2020] note that not exercising skepticism often "offers a high probability of good performance on the objective budget measure, and the low base rate of misstatements means that the likelihood of missing a misstatement is relatively low." Further, recent research indicates that client management are not apt to have explanations for red flags identified by analytic tools and are also hesitant to pay additional fees related to auditor use of such tools (Austin et al. [2019]). Thus, when the rate of false positives is higher, the risk of costly skepticism increases and the benefit of consistently rewarding costly skepticism is likely to diminish as auditors regress back to the risk-

averse, "quit while you're ahead" mindset described by Brazel et al. [2020]. As such, we predict that the positive effect of consistent rewards for costly skepticism will diminish when false positive rates are higher (vs. lower). Formally stated,

H3: The positive relation between consistently rewarding costly skepticism and skeptical actions will be reduced when there is a higher rate of false positives.

II. METHOD

Participants

We tested our hypotheses via an experiment. The participants in the experiment were 113 audit seniors from two Big Four accounting firms and a national firm. The participants completed the experiment during firm-sponsored training sessions. Their task was a substantive analytical procedure for sales, which is appropriate for audit seniors (Trompeter and Wright [2010]). The mean task completion time was 27 minutes.¹⁴ On average, participants had 49 months of experience and had conducted substantive analytical procedures related to sales three times during their careers. Our experiment was a 2 (Consistent Reward: *present* or *absent*) × 2 (False Positive Rate: *lower* or *higher*) between-participants design. We randomly assigned participants to one of the four conditions.^{15,16}

Description of the Experimental Context and Independent Variables

The experimental materials first informed participants that they were the audit senior conducting year-end procedures for the audit of Madison, Inc., a publicly traded manufacturing company. Given our focus on reward consistency, participants were told that this was their third

¹⁴ Given the notion that the use of data analytics may eventually make audits more efficient (Austin et al. 2019), it is interesting to note that our average completion time (27 minutes) is substantially lower than the completion time of the participants in Brazel et al. [2020] who completed a substantive analytic procedure with similar data, but were not provided with visualizations from an analytic tool (34 minutes).

¹⁵ We obtained Institutional Review Board approval for the experiment used in this study.

¹⁶ There are no differences across conditions for any demographic or experienced-based measures, such as general auditing experience, industry experience, experience with analytical procedures, and perceptions of the usefulness of audit data analytics.

year serving on the Madison engagement. Participants then read that they had engaged in costly skepticism while conducting interim procedures. Specifically, while testing depreciation expense for property, plant, and equipment (PP&E), they noted year-to-year changes in the client's estimates for the useful lives of several asset classes. Although the client provided supporting internal documentation, participants had deemed it appropriate to exercise heightened skepticism and conducted additional procedures to verify the changes. However, the additional procedures identified no misstatement, caused friction with the client, and caused the audit of PP&E to go over budget. The case then provided participants with their manager's evaluation of their performance for this interim testing (the *REWARD* manipulation). Participants in the consistent *REWARD* present condition were told the following:

You received an evaluation from your audit manager related to your interim testing at Madison. Your audit manager specifically identified the useful life investigation in the evaluation and considered the investigation appropriate. Your audit manager evaluated your performance during interim testing as *Exceeded Expectations*.

This evaluation *is consistent* with the evaluations you have received in the past from your manager when you investigated evidence inconsistencies, exceeded the budget and/or strained relations with management, and found a reasonable explanation for the evidence inconsistencies (e.g., false positives).

Participants in the consistent *REWARD* absent condition were told the following:

You received an evaluation from your audit manager related to your interim testing at Madison. Your audit manager specifically identified the useful life investigation in the evaluation and considered the investigation appropriate. Your audit manager evaluated your performance during interim testing as *Met Expectations*.

This evaluation *is consistent* with the evaluations you have received in the past from your manager when you investigated evidence inconsistencies, exceeded the budget and/or strained relations with management, and found a reasonable explanation for the evidence inconsistencies (e.g., false positives).¹⁷

¹⁷ To be clear, in both *REWARD* conditions the presence or absence of the reward at interim was *consistent* with past evaluations received from the supervisor. Therefore, we do not manipulate *consistency* in this study. As a manipulation check for *REWARD*, a post-experimental question asked participants: "Based on the information provided for the audit of Madison, describe how *consistently* your *audit manager* at Madison *would reward* you for investigating evidence inconsistencies that caused budget overages and/or strain relations with management, but resulted in a reasonable

Next, participants completed a year-end substantive analytical procedure related to the sales account for one of Madison's operating units, Madison Sporting Goods, which reported positive sales growth. Specifically, their task was to develop an expectation for Madison's Sporting Goods Sales account, compare the expectation to the recorded balance, and conclude on whether additional testing was required for the account (our dependent variable *PS* described below). The case included a rich set of financial and non-financial data from the sources suggested by AS 2305 (PCAOB [2019b]) (i.e., prior year balances, budgets, industry growth rates, growth in related accounts, and growth in non-financial measures). All *financial* measure trends were positive and consistent with the client's current year sales growth of 9%. The trend for all *non-financial* measures (NFMs), such as the number of customer accounts, patents, and employees, was negative and *inconsistent* with reported sales. The NFMs provided to all participants exhibited declines of approximately 21%, reflecting the NFM red flag observed by Brazel et al. [2009] for fraud firms (i.e., a 30% difference between sales growth and growth in related NFMs).¹⁸

The financial and nonfinancial measures were presented to participants in three stages. To offer an example calculation/expectation, we first presented all participants with prior year balances and industry sales trends and a calculation combining the two sources that provided an

explanation for the evidence inconsistencies (e.g., false positives)." The response scale ranged from 1 (Not Consistently Reward) to 10 (Consistently Reward). The mean response in the consistent *REWARD* present condition (7.07) was significantly higher than the mean for the consistent *REWARD* absent condition (3.87, p < .001).

¹⁸ Recent research indicates that the use of data analytics often spurs the incorporation of NFMs into audit testing (Rose et al. 2017; Austin et al. 2019). With respect to whether it was appropriate for the auditors in our study to further investigate the NFM red flag, Brazel, Jones, and Prawitt [2014] "surveyed 23 audit managers and partners (mean experience =16.5 years) and asked the following: Please rate the effect of sales growth outpacing NFM growth by 25 percent on the overall assessment of fraud risk. These professionals responded on an 11-point scale where: 1 = "definitely reduce fraud risk," and 11 = "definitely increase fraud risk." The mean response was 9.43. These partners and managers rated the inconsistency as significantly higher than several other common red flags (e.g., high accruals, CFO turnover in the current year, management being extremely reluctant to post any audit adjustments (p-values < 0.05))." In addition, audit senior participants in Brazel et al. [2016] indicated a strong belief that the NFM red flag should be investigated.

expectation consistent with Madison Sporting Goods Sales.¹⁹ Second, participants were told that they had used a data analytic tool (Tableau) to create visuals comparing Madison Sporting Goods' Sales over the last five years to various sources of financial and non-financial data.

Prior to the presentation of the visuals, participants were informed of the false positive rate for outliers/inconsistencies identified via data analytic tools in the past for Madison (the *FALSE POSITIVE* manipulation). Participants in the lower *FALSE POSITIVE* condition were told the following:

Approximately 25% of outliers/inconsistencies identified via data analytic tools in the past for Madison have NOT identified misstatements (false positives).

Participants in the higher FALSE POSITIVE condition were told the following:

Approximately 75% of outliers/inconsistencies identified via data analytic tools in the past for Madison have NOT identified misstatements (false positives).²⁰

While we are not aware of any prior literature that documents "lower" and "higher" false positive rates for audit data analytics, our chosen rates were influenced by discussions with practitioners. Through these discussions we became informed about how such rates vary in practice and the importance of interpreting a false positive rate in light of its "hit rate." Specifically, given evidence suggesting that the base rate of misstated accounts is quite low (e.g., Durney et al. [2014], Brazel et al. [2020]), practitioners viewed our higher false positive rate of 75% as

¹⁹ Brazel et al. [2014] illustrate that auditors are apt to use prior year balances and industry data over other sources while performing substantive analytical procedures for sales. We also provided the example calculation as means of measuring whether the participant *did not* incorporate the visualizations/data analytics that followed into their expectation (and to illustrate that, as in practice, the participant was not required to use the visualizations when developing their expectation). Analysis of our experimental data indicate that neither of our study's independent variables significantly affected whether participants chose to use the provided expectation (p's > .05). In the section titled The Link Between Skeptical Judgments and Skeptical Actions we further examine our participants' skeptical judgments (i.e., evidence evaluation) and their correlations with skeptical actions (our key dependent variable *PS*).

²⁰ As a manipulation check for *FALSE POSITIVE*, a post-experimental question asked participants: "For the audit of Madison, please describe the *rate* at which outliers/inconsistencies identified via data analytic tools have *NOT identified a misstatement (false positives).*" The response scale ranged from 1 (Very Low) to 10 (Very High). The mean response in the lower *FALSE POSITIVE* condition (3.27) was significantly lower than the mean for the higher *FALSE POSITIVE* condition (7.14, p < .001). Also, these mean responses indicate that neither our lower nor higher *FALSE POSITIVE* conditions were perceived as extreme by our participants.

indicative of well calibrated analytic tools (25% of exceptions identified by the analytic tools are eventually determined to be a misstatement (i.e., a 25% hit rate)). Practitioners generally viewed our lower false positive rate of 25% as a *very* well calibrated set of analytics (75% of exceptions identified by the tools are determined to be as a result of a misstatement). Further, our experimental instrument was reviewed for realism and clarity by several partners of one of the participating Big Four firms. As part of this process, we specifically requested the partners to inform us if our lower and higher false positive manipulations were not in line with what they had observed in practice. The reviewing partners did not provide any adjustments to our lower and higher false positive rates. In sum, we believe that both our lower and higher false positive conditions reflect settings that auditors would expect to encounter when using well calibrated analytic tools in the field.

The visualizations were developed in Tableau (an analytic tool used by auditors) and are presented in Figure 1. Note that Visualizations 1-3 and 5 all present consistent relations between Madison Sporting Goods Sales and related financial measures (e.g., prior year balances). However, Visualization 4 highlights the NFM red flag being present in the current year (i.e., a substantial inconsistency between sales growth and related NFM growth). Last, a table was presented to participants that provided the underlying prior year and current data used in the aforementioned visualizations.

Dependent Variable

After calculating an expectation for sales and computing the difference between their expectation and reported sales, participants provided the study's primary dependent variable: *PROFESSIONAL SKEPTICISM (PS)*. Specifically, participants chose one of two conclusions about whether or not the sales account warranted additional investigation: (1) "The balance appears reasonable. No additional work would be required related to this analytical procedure" or (2)

"Additional work would be required related to this analytical procedure before concluding the balance appears reasonable." We coded *PS* as "1" if participants chose option (2), and "0" if participants chose option (1). Given the presence of the NFM red flag/inconsistent evidence, coding further investigation as indicative of an appropriate skeptical action is consistent with the professional skepticism literature (e.g., IAASB 2004; Nelson 2009; Brazel et al. 2016).

IV. RESULTS

Tests of H1 and H2

We manipulated two independent variables (*FALSE POSITIVE* and *REWARD*) and auditor participants provided a conclusion to their testing that we classified as a skeptical action (*PS*) to test our predictions. Table 1 Panel A shows mean *PS* across the *FALSE POSITIVE* and *REWARD* conditions. We first observe a negative relationship between *PS* and *FALSE POSITIVE*, which is consistent with our prediction in H1. Specifically, we find that, when red flags are identified using a data analytic tool associated with a higher rate of false positives, skeptical actions diminish (0.49 versus 0.34, p = 0.052, one-tailed). We also observe a positive relationship between *PS* and *REWARD*, which aligns with our prediction in H2. Consistently rewarding auditors for engaging in costly skepticism increases skeptical actions in subsequent tasks (0.29 versus 0.53, p = 0.004, one-tailed).

We formally test our H1 and H2 using logistic regression. Table 1 Panel B presents the results of the model including the two independent variables to test the main effects of each variable on *PS*. Supporting both H1 and H2, we observe a significant and negative main effect for *FALSE POSITIVE* ($\beta = -0.685$, p = 0.044) and a significant and positive main effect for *REWARD* ($\beta = 1.064$, p = 0.004). As noted above, auditors are less likely to exercise skepticism when *FALSE POSITIVE* is higher and are more likely to exercise skepticism when *REWARD* is present.

Specifically, we find that auditors are 33% *less* likely to exercise skepticism when *FALSE POSITIVE* is higher (vs. lower), and auditors are 74% *more* likely to exercise skepticism when *REWARD* is present (vs. absent).²¹ However, it is important to note that these main effects should be interpreted with caution as the model in Panel B does not consider whether the effect of *REWARD* is conditional upon the *FALSE POSITIVE* rate (i.e., the interaction posited by H3).

Tests of H3

We formally test H3 by re-performing our logistic regression with the interaction between *FALSE POSITIVE* and *REWARD* included. We also test the specific predicted pattern of H3 with planned comparisons.²² As posited in H3 and illustrated in Figure 2, the effect of *REWARD* on *PS* diminishes when the rate of false positives goes from lower to higher. The interaction term in Panel C of Table 1 is significant ($\beta = -1.342$, *p* < .05). When *REWARD* is present and *FALSE POSITIVE* is low, we observe an 84% increase in the probability of exercising skepticism, however, when *FALSE POSITIVE* is high, the effect of *REWARD* on PS is reduced to only a 59% increase.²³ It is important to note that the main effects for both *REWARD* and *FALSE POSITIVES* in *this model* only present the residual effect of each variable after the interaction has been accounted for, and therefore, it would not be appropriate to test H1 and H2 with the model in Panel C.

We further examine the specific pattern of results predicted by H3 with planned comparisons in Panel D of Table 1. We find a significant effect of *REWARD* on *PS* when the data analytic tools used on the audit had a lower *FALSE POSITIVE* rate (0.69 vs. 0.29, p < .001, see

²¹ Logistic regression results present the odds of exercising skepticism (PS = 1) over not exercising skepticism (PS = 0). We calculate the percent change in the probability of exercising skepticism as EXP(β)/(1+ EXP(β)). For *FALSE POSITIVE* we calculate EXP(-.685)/(1+EXP(-.685)) and for REWARD we calculate EXP(1.064)/(1+EXP(1.064)).

²² We do not use contrast coding to test the predicted pattern because contrast coding examines the linear relation among means while our DV is binary, and thus, non-linear. Instead, we utilize planned comparisons to evaluate predicted differences in cell means.

²³ The effect of *REWARD* when *FALSE POSITIVE* is lower is calculated as EXP(1.715)/(1+EXP(1.715)). The effect of *REWARD* when *FALSE POSITIVE* is higher is calculated as EXP(-1.342+1.715)/(1+EXP(-1.342+1.715)).

Table 1 Panels A and D), but not in the higher *FALSE POSITIVE* setting. Results support our prediction in H3 that the positive effect of *REWARD* on *PS* would diminish when data analytic tools had a history of higher *FALSE POSITIVE* rates. Importantly, when *REWARD* is absent, as is the norm in most audit settings (Brazel et al. [2016], Brazel et al. [2019], Brown and Millar [2020]), we see consistently low *PS* levels (.29 and .30, Panel A). When *REWARD* is not present, auditor *PS* is low regardless of the *FALSE POSITIVE* rate.

Additional Analyses

The Effect of Message Severity on Skeptical Action

Recall that in the development of H1 we note that, when the false positive rate is higher, the perceived message severity is likely to be low. Because they have experienced a higher rate of false positives in the past, auditors will perceive a lower risk associated with dismissing red flags identified by analytic tools and will be less apt to rely on the analytic and incorporate the inconsistent NFMs into their testing. In turn, professional skepticism will be lower. In Figure 3 we examine whether message severity mediates the effect of *FALSE POSITIVE* on *PS*. We measure *MESSAGE SEVERITY* as the average of two measures. The first measure is the extent to which participants reported to rely on the visuals from the data analytic tool (*RELY*) on a scale ranging from 1 (No Reliance) to 10 (Heavy Reliance). The second measure is participants' rank of the importance of the non-financial measures (NFMs) in the development of their expectation for the current year's sales, where a rank of 1 is perceived as least important and a rank of 5 is perceived as most important (*NFM RANK*).²⁴ Results illustrated in Figure 3 show that *MESSAGE SEVERITY* mediates the relationship between *FALSE POSITIVE* and *PS* (95% two-tailed CI: - 1.04, -0.18).

 $^{^{24}}$ Participants rated all five sources of data provided (e.g., prior period financial data, industry data, NFMs) with 1 = "most important" and 5 = "least important. We then reverse coded participant responses.

Specifically, we find that when *FALSE POSITIVE* is higher, auditors perceived the signal provided by the analytic as less severe (b = -1.13, p < 0.01) when developing their expectation for sales (indicative of auditors dismissing the NFM red flag highlighted by the analytic). This lower perception of *MESSAGE SEVERITY*, in turn, leads to lower *PS* (b = 0.46, p < 0.01).^{25,26}

The Link Between Skeptical Judgments and Skeptical Actions

Nelson [2009] describes a model of *PS* that suggests it includes two related components skeptical *judgments* and skeptical *actions*—and the former should influence the latter (see also Hurtt et al. [2013]). Consistent with Nelson [2009] and Brazel et al. [2020], our primary measure of *PS* represents a skeptical action or the decision to further test an account where an evidence inconsistency exists. Related to skeptical *judgment* we measured the participants' recognition of the NFM red flag (*RECOGNITION*). See the Appendix for variable descriptions.

Despite false positive rates and rewards impacting our auditors' skeptical actions, we do not observe any variation in their skeptical judgments. In Panel A of Table 2 we find no difference in *RECOGNITION* of the NFM red flag across all conditions. All correlations with *RECOGNITION* are not significant in Panel B (see also the ANOVA results in Table 2, Panel C). Auditors in all experimental conditions typically recognized the red flag, yet unless *both* false positives were lower

²⁵ We do not observe that *REWARD* moderates the effect of *FALSE POSITIVE* illustrated in Figure 3.

²⁶ We also consider the effects of message ambiguity on skeptical actions. We use *RECOGNITION* and *RISK OF MATERIAL MISSTATEMENT (RMM)* as our measures of message ambiguity as the more ambiguous the message, the less likely it is to be recognized as a red flag and indicative of a higher risk of material misstatement (see the APPENDIX for variable descriptions). We do not find a significant mediating effect of *RECOGNITION* or *RMM* on the relationship between *FALSE POSITIVES* and *PS. RECOGNITION* is consistent across all conditions (see next section for a further discussion). We do note that *RMM* mediates the effect of *REWARD* on *PS* (the relationship predicted by H2) such that auditors assess *RMM* higher when *REWARD* is present and are therefore more likely to exercise professional skepticism. Our experiment included the presence of a fraud red flag in all experimental conditions. Thus, better calibrated risk assessments could be another positive by-product of rewarding costly skepticism. On the other hand, rewards positively impacting risk assessments could be an unintended consequence if such a relation is observed in low risk settings. The links between rewards, risk assessments, and over- and under-auditing appears to be a fruitful area for further research.

and rewards were present, our auditors' skeptical judgments did not typically convert to skeptical actions (i.e., investigation of the red flag).

To Document, or Not Document: That is the Question

As noted above and illustrated in Panel A of Table 2, we observe that *RECOGNITION* of the NFM red flag is relatively high and consistent across all four conditions, while *PS* varies between conditions. Indeed, in Panel B we do not observe a significant correlation between *RECOGNITION* and *PS*. Given these inconsistent findings between *RECOGNITION* and *PS*, the question that remains is whether participants chose to document the NFMs when calculating their estimates of sales. For our three conditions where *PS* is low, *not documenting* the NFMs could perhaps be a strategic decision in case the audit is inspected or brought into litigation (i.e., the auditor does not want to document that they identified, but failed to question, the red flag (e.g., Reffett [2010]). However, not documenting the NFMs also reduces the likelihood that the issue is identified by a superior or reviewer during the review process, which compromises the review process as a quality control mechanism.

To examine this question we reviewed the calculations that participants documented in their workpaper to support their expectation of sales. We created a variable which is coded 1 if the participant documented the NFMs and 0 otherwise (*DOCUMENTATION*). We observe in Panel A of Table 2 that, when *FASLE POSITIVE* is higher, participants typically chose to *not* document the NFMs (.19 and .17). In Panel D will illustrate significant negative relations between *FALSE POSITVE* and *DOCUMENTATION* (p's < 0.01). When *FASLE POSITIVE* is higher, auditors appear to cover their tracks and not document the NFM red flag that they recognize, but do not intend to investigate. As would be expected, when PS is highest (.69 in the lower FALSE POSITIVE/REWARD present condition), we observe a higher DOCUMENTATION rate (.45). Unexpectedly, when FALSE POSITIVE is lower and REWARD is absent we observe a low level of PS (.29), but half the auditors document the NFMs (.50). Perhaps, given the lower FALSE POSITIVE rate, auditors were more comfortable in documenting the NFMs and potentially investigating the red flag if questioned by their supervisor/reviewer. These auditors may also be "passing the buck" of potentially costly skepticism to their supervisor who has not consistently rewarded the behavior. Overall, DOCUMENTATION appears to be crucial to PS as we illustrate in Panel E that DOCUMENTATION is significantly correlated with PS in every experimental condition. Given the tendency for analytics to identify anomalies, how and why auditors choose to document or not document inconsistent evidence derived from analytic tools is a fruitful area for further research.

The Role of Auditor Risk-Aversion

In the development of H3 we note that when the rate of false positives is higher, the risk of costly skepticism increases and the benefit of consistently rewarding costly skepticism is likely to diminish as auditors regress back to a risk-averse mindset. While auditors tend to be risk-averse (Lennox and Kausar [2017]), this trait likely varies between auditors and some auditors may not typically exhibit risk-aversion. As such, if risk-aversion is playing a role in our H3 result, we would not expect H3 (or H1) to be supported when examining auditors who are less risk-averse.

To test this conjecture, we measured our participants' *RISK-AVERSION* levels (see the APPENDIX for a description) and examined a moderated moderation model. In Figure 4 we illustrate that the effect of *REWARD* on *PS* is moderated by the *FALSE POSITIVE* rate (consistent with H3), but the extent of the moderation is dependent upon the auditor's level of *RISK*-

AVERSION (χ^2 =9.98, p < 0.01).²⁷ Specifically, the *FALSE POSITIVE* rate moderates the effect of *REWARDS* when participant *RISK-AVERSION* levels are higher, but not when *RISK-AVERSION* is lower. Further, as noted in our reporting of conditional effects in Figure 4, when *FALSE POSITIVE* rates are higher, *REWARDS* continue to positively affect *PS* for auditors with lower *RISK-AVERSION* (see the bolded first row, p < .01). The interaction predicted by H3 does not occur for auditors with lower levels of *RISK-AVERSION*, as these individuals do not have the risk-averse mentality to revert back to when the *FALSE POSITIVE* rate is higher.²⁸

V. CONCLUSION

Data analytics are designed to help auditors better understand their client's financial and nonfinancial data by identifying patterns and relationships and detecting anomalies or red flags within the data (AICPA [2017]). One concern over the use of data analytics is the presence of false positives, or the extent to which data analytic tools identify anomalies or red flags that, after investigation, are reasonable and explained variations in the data (AICPA [2017], Johnson and Wiley [2019], Brazel et al. [2020]). The objective of this study was to investigate if varying rates of false positives impact auditor skepticism toward red flags identified by data analytics tools. We also examined the extent to which consistent rewards for skepticism can improve the application of skepticism on audits employing data analytics.

Using an experiment with practicing auditors where a red flag is identified by a well calibrated analytic, we observe that when false positive rates are higher, skepticism levels are low.

²⁷ We also test a simple moderation model in which *FALSE POSITIVE* moderates the effect of *REWARD* on *PS* (H3) and find a significant moderating effect (b = -1.34, p < 0.05).

²⁸ We expect those with higher *RISK-AVERSION* levels to "revert" back to the safe choice of not investigating when *FALSE POSITIVE* is higher, but those with lower levels of risk-aversion have little such mentality to revert back to. Consistent with this prediction, we find that auditors with lower levels of *RISK-AVERSION* are not affected by the false positive rate as predicted by H1 (β = -1.844, p = 0.152). We examine this by re-performing our tests of H1 with only those participants identified as low on *RISK-AVERSION* (-1 standard deviation from the mean).

We find that consistent rewards for costly skepticism (appropriate skepticism that incurs costs, but does not identify a misstatement) cause auditors to more appropriately apply skepticism. When analytic tools are very well calibrated and the rate of false positives is lower, we find that consistent rewards significantly improve the skepticism of our auditors. Unfortunately, prior research has illustrated that such consistent rewards are rare (e.g., Brazel et al. [2016]). We observe low levels of skepticism when consistent rewards are absent, even when the false positive rate is lower. As such, our findings suggest that a culture change of consistently rewarding costly skepticism is needed in order for audit firms to effectively use analytic tools to enhance audit quality.

We also illustrate that the positive effects of rewards are limited in that we do not see improvements in skepticism when the false positive rate is higher, and it is probable that any additional investigation will lead to costly skepticism. Therefore, in such settings, firms should examine if the false positive rate is or is not out of their control and determine the steps that can be taken to reduce the rate. As described by the AICPA [2017] and Johnson and Wiley [2019], a higher false positive rate could be due to the analytic not being properly planned and performed or the tool not being a good fit for the audit test. On the other hand, a higher rate could be as a result of a client with operations and accounting processes that are prone to exhibit exceptions. For example, a client may be more apt to have cash collections from customers that exceed or fall below related billings (AICPA [2017]). A company could also be more prone to experience time delays between investments in operations/NFMs and their related impact on the financial statements (Brazel and Lail [2019]). Under such settings, engagement teams can employ grouping and filtering techniques to their clients' data (AICPA [2017]; Johnson and Wiley [2019]; No et al. [2019]). Examining such ways to reduce false positive rates represents a fruitful area for future research.

Since the positive effects of rewards diminish when false positive rates rise, further research can also examine other potential supervisor interventions that may improve auditor use of analytics in more ambiguous settings. For example, more explicit discussions about when further investigation is expected, if and how exceptions can be dismissed, and when consultation with the supervisor is appropriate might improve auditors reactions to anomalies identified by analytic tools. Last, recent research has discussed the use of analytics by corporate management involved in the financial reporting process (Austin et al. [2019]). Examining the impediments to managers' use of analytic tools and what factors influence their effective application would be an interesting course for future research.

REFERENCES

- Agoglia, C., R. Hatfield, and T. Lambert. "Audit Team Time Reporting: An Agency Theory Perspective." *Accounting, Organizations and Society* 44 (2015): 1 14.
- Anderson, S. B., J. L. Hobson, and M.E. Peecher. "The Joint Effects of Rich Data Visualization and Audit Procedure Categorization on Auditor Judgment." Working paper, University of Illinois at Urbana-Champaign, 2020.
- Arens, A. A., R. J. Elder, M. S. Beasley, and C. E. Hogan. Auditing and Assurance Services: Auditing and Assurance Services. 17th ed. New York City: Pearson Education, Inc, 2020.
- American Institute of Certified Public Accountants (AICPA). *Guide to Data Analytics*. Durham: AICPA, 2020.
- Austin, A. A., T. Carpenter, M. H. Christ, and C. Nielson. "The Data Analytics Transformation: Evidence from Auditors, CFOs, and Standard-Setters." Working paper, University of Richmond, 2019.
- Baadder, G., and H. Krcmar. "Reducing False Positives in Fraud Detection: Combining the Red Flag Approach with Process Mining." *International Journal of Accounting Information Systems* 31 (2018): 1-16.
- Barr-Pulliam, D. "The Effect of Continuous Auditing and Role Duality on the Incidence and Likelihood of Reporting Management Opportunism." *Management Accounting Research* 44 (2019): 44-56.
- Barr-Pulliam, D., H. Brown-Liburd, and K. Sanderson. "The Effects of the Internal Control Opinion and use of Audit Data analytics on Perceptions of Audit Quality, Assurance, and Auditor Negligence." Working paper, University of Louisville, 2019.
- Brazel, J. F., K. L. Jones, and M. F. Zimbelman. "Using Nonfinancial Measures to Assess Fraud Risk." *Journal of Accounting Research* 47 (2009): 1135–1166.
- Brazel, J. F., K. L. Jones, and D. F. Prawitt. "Auditors' Reactions to Inconsistencies between Financial and Nonfinancial Measures: The Interactive Effects of Fraud Risk Assessment and a decision Prompt." *Behavioral Research in Accounting* 26 (2014): 131-156.
- Brazel, J. F., S. B. Jackson, T. J. Schaefer, and B. W. Stewart. "The Outcome Effect and Professional Skepticism." *The Accounting Review* 91 (2016): 1577 1599.
- Brazel, J. F., C. Gimbar, E. Maskymov, and T. J. Schaefer. "The Outcome Effect and Professional Skepticism: A Replication And A Failed Attempt At Mitigation." *Behavioral Research in Accounting* 31 (2019): 135-143.
- Brazel, J. F., and B. Lail. "How the interplay between financial and non-financial measures affects management forecasting behavior." *Journal of Management Accounting Research* 31 (2019): 41-63.
- Brazel, J. F., J. Leiby, and T. Schaefer. "Do Rewards Encourage Professional Skepticism?" Working paper, North Carolina State University, 2020.
- Brown, O., and M. I. Millar. "A Matter of Perspective: Mitigating the Outcome Effect in Auditor Performance Evaluations." Working paper, Baylor University, 2020.
- Brown-Liburd, H., H. Issa, and D. R. Lombardi. "Behavioral implications of big data's impact on audit judgment and decision making and future research directions." *Accounting Horizons* 29 (2015): 451-468.
- Cao, M., R. Chychyla, and T. Stewart. "Big data analytics in financial statement audits." *Auditing: A Journal of Practice & Theory* 29 (2015): 423-429.

- Davidson, R. A., and J. T. Dalby. "Personality profiles of Canadian public accountants." *International Journal of Selection and Assessment* 1 (1993): 107–116.
- Davidson, B. I., N. Desai, and G. J. Gerard.. "The effect of continuous auditing on the relationship between internal audit sourcing and the external auditor's reliance on the internal audit function." *Journal of Information Systems* 27 (2013): 41-59.
- Davis, J. M, and B. M. Tuttle. "A heuristic-systematic model of end-user information processing when encountering IS exceptions." *Information and Management* 50 (2013): 125-133.
- Deloitte. *Quality and Innovation: Enhancing the Audit*. 2016 Available at: https://www2.deloitte.com/vi/en/pages/about-deloitte/articles/gr14-audit.html.
- Durney, M., R. Elder, and S. Glover. "Field Data on Accounting Error Rates and Audit Sampling." *Auditing: A Journal of Practice & Theory* 33 (2014): 79 110.
- Ernst & Young (EY). *Big Data: Changing the Way Businesses Complete and Operate*. 2014. Available at: https://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/%24FILE/EY-Insights-on-GRC-Bigdata.pdf.
- Farmer, T. A. "Testing the Effect Of Risk Attitude on Auditor Judgment using Multiattribute Utility Theory." *Journal of Accounting, Auditing & Finance* 8 (1993): 91–110.
- Huang, W., P. Eades, and S. Hong. "Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective." *Information Visualization* 8 (2009): 139-152.
- Hurtt, R. K., H. Brown-Liburd, C. E. Earley, and G. Krishnamoorthy. "Research on Auditor Professional Skepticism: Literature Synthesis and Opportunities for Future Research." *Auditing: A Journal of Practice & Theory* 32 (2013): 45-97.
- International Auditing and Assurance Standards Board (IAASB). *Objective and General Principles Governing an Audit of Financial Statements*. International Standard on Auditing No. 200. New York, NY: International Federation of Accountants, 2004.
- International Auditing and Assurance Standards Board (IAASB). Invitation to Comment: Enhancing Audit Quality in the Public Interest: A Focus on Professional Skepticism, Quality Control and Group Audits. New York, NY: IFAC, 2015.
- Johnson, R., and L. Wiley. *Auditing: A Practical Approach with Data Analytics*. 1st ed. Hoboken: John Wiley and Sons, Inc, 2019.
- Kogan, A., B. Mayhew, and M. Vasarhelyi. "Audit Data Analytics Research-An Application of Design Science Methodology." *Accounting Horizons* 33 (2019): 69-73.
- KPMG. *Data & Analytics: Unlocking the Value of Audit.* 2014. Available at: https://home.kpmg/xx/en/home/insights/2015/02/audit-data-analytics-unlocking-value-of-audit.html.
- KPMG. 2016 KPMG International Transparency Report Our Relentless Focus on Quality 2016. Available at: https://home.kpmg/xx/en/home/campaigns/2016/12/kpmg-international-transparency-report-2016.html.
- Krahel, J. P., and W. R. Titera. "Consequences of Big Data and Formalization on Accounting and Auditing Standards." *Accounting Horizons* 29 (2015.): 409-422.
- Lennox, C. S., and A. Kausar. "Estimation Risk and Auditor Conservatism." *Review of Accounting Studies* 22 (2017): 185-216.
- Li, P., D. Y. Chan, and A. Kogan. "Exception Prioritization in Continuous Auditing: A Framework and Experimental Evaluation." *Journal of Information Systems* 30 (2016):135-157.

- Luippold, B. L., and T. E. Kida. "The Impact of Initial Information Ambiguity on the Accuracy of Analytical Review Judgments." *Auditing: A Journal of Practice & Theory* 31 (2012): 113-129.
- Nelson, M. "A Model and Literature Review of Professional Skepticism in Auditing." *Auditing: A Journal of Practice & Theory* 28 (2009): 1-34.
- No, W., K. Lee, F. Huang, and Q. Li. "Multidimensional Audit Data Selection (MADS): A Framework for Using Data Analytics in the Audit Data Selection Process." *Accounting Horizons* 33 (2019): 127-139.
- Nolder, C. J., and K. Kadous. "Grounding the Professional Skepticism Mindset And Attitude Theory: A Way Forward." *Accounting, Organizations and Society* 67 (2018): 1-14.
- Prendergast, C. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1999): 7-63.
- PricewaterhouseCoopers (PwC). 2014 State of the Internal Audit Profession Study. 2014. Available at: <u>http://www.pwc.com/en_M1/m1/publications/documents/pwc-state-of-the-internal-audit-profession-2014.pdf</u>.
- PwC. *PwC Transforms Audit Practice with Data Analytics*. 2015. Available at: https://www.accountingtoday.com/opinion/pwc-transforms-audit-practice-with-data-analytics.
- Protiviti. 2014 Internal Audit Capabilities and Needs Survey. 2014. Available at: http://www.protiviti.com/en-US/Pages/IA-Capabilities-and-Needs-Survey.aspx.
- Public Company Accounting Oversight Board (PCAOB). *Maintaining and Applying Professional Skepticism in Audits*. Staff Audit Practice Alert No. 10. Washington, D.C.: PCAOB, 2012.
- Public Company Accounting Oversight Board (PCAOB). *Due Professional Care in the Performance of Work*. Auditing Standard 1015. Washington, DC: PCAOB, 2019a
- Public Company Accounting Oversight Board (PCAOB). Substantive Analytical Procedures. Auditing Standard 2305. Washington, DC: PCAOB, 2019b
- Reffett, A. B. "Can Identifying and Investigating Fraud Risks Increase Auditor Liability?" *The Accounting Review* 85 (2010): 2145-2167.
- Rose, A., J. Rose, K. Sanderson, and J. Thibodeau. "Where Should Audit Firms Introduce Analyses of Big Data into the Audit Process?" *Journal of Information Sciences* 31 (2017): 81-99.
- Schneider, G. P., J. Dai, D. J. Janvrin, K. Ajayi, and R. L. Raschke. "Infer, Predict, and Assure: Accounting Opportunities in Data Analytics." *Accounting Horizons* 29 (2015.): 719-742.
- Trompeter, G., and A. Wright. "The World has Changed—Have Analytical Procedure Practices?" *Contemporary Accounting Research* 27 (2010): 669–700.
- Tversky, A., and D. Kahneman. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (1981): 453 458.
- Vessey, I. "Cognitive Fit: A Theory Based Analysis of the Graphs Versus Tables Literature." *Decision Sciences* 22 (1991): 219-240.
- Walker, K., H. Brown-Liburd, and A. Lewis. "The emergence of data analytics in auditing: Perspectives from internal and external auditors through the lens of institutional theory." Working paper, University of Wisconsin – Madison, 2019.
- Wang, T., and R. Cuthbertson. "Eight Issues on Audit Data Analytics We Would Like Researched." *Journal of Information Systems* 29 (2015): 155–162.

Yoon, K., L. Hoogduin, and L. Zhang. "Big Data as Complementary Audit Evidence." Accounting Horizons 29 (2015): 431-438.

APPENDIX VARIABLE DESCRIPTIONS

Variable	Description
REWARD	Manipulated between-subjects as a history of consistent rewards for
	engaging in costly skepticism by the auditor manager (present or absent).
FALSE POSITIVE	Manipulated between-subjects as either a 25% (Lower) or 75% (Higher) rate
	of false positives generated by the data analytic tool.
PROFESSIONAL	Assessments about whether the account warrants more investigation and
SKEPTICISM (PS)	coded as $0 =$ "Requires no more work and I conclude it is reasonable" or 1
	= "Requires more work before I can conclude it is reasonable."
MESSAGE SEVERITY	Average of <i>RELY</i> and <i>NFM RANK</i> where higher scores equal higher message severity.
NFM RANK	Participants' rank of the importance of the non-financial measures (NFMs) on a scale from 1 (Least Important) to 5 (Most Important).
RELY	Participants were asked to describe the extent to which they relied on the
	<i>visuals</i> from the data analytic tool on a scale ranging from 1 (No Reliance)
	to 10 (Heavy Reliance).
RECOGNITION	Participants were asked to describe the <i>difference or inconsistency</i> between
	the growth in the Madison Sporting Goods Sales account and the growth in
	related non-financial measures (e.g., number of employees, square footage
	of production space, etc.) on a scale ranging from 1 (Very Small) to 10 (Very
	Large).
RISK OF MATERIAL	Participants were asked their assessment of the likelihood that the Madison
MISTATEMENT (RMM)	Sporting Goods Sales account was materially misstated on a scale ranging
	from I (Low) to 10 (High).
DOCUMENTATION	Measures whether participants documented NFMs in their calculations of
	their estimates of current year Madison Sporting Goods Sales and coded as
	0 = did not document NFMs or 1 = did document NFMs.
RISK-AVERSION	Measured with three post-test questions measuring preferences between (1)
	a choice that guarantees an average outcome or (2) a choice that has
	uncertainty about whether it will produce an excellent outcome, an average
	outcome, or a poor outcome. Participants assess their preferences on a scale
	from 0 to 100, with $0 =$ choice 1, $50 =$ indifferent, and $100 =$ choice 2. We
	three tradeoffers that higher values represent higher risk eversion (Drozel
	at al [2020])
	et al. [2020]).



FIGURE 1: Visualizations Used in the Experiment





FIGURE 2: Cell Means for Professional Skepticism (*PS*) by *REWARD* and *FALSE POSITIVE*



The dependent variable (*PS*) measures the auditors' level of professional skepticism. PS is the auditors' assessment about whether the account warrants more investigation and coded as 0 = "Requires no more work and I conclude it is reasonable" or 1 = "Requires more work before I can conclude it is reasonable."

The conditions are *REWARD* (manipulated between-subjects as a history of consistent rewards for engaging in costly skepticism by the auditor manager; absent or present) and *FALSE POSITIVE* (manipulated between-subjects as either a 25% (*Lower*) or 75% (*Higher*) rate of false positives generated by the data analytic tool.)



Figure 3 shows the results (coefficients) of a mediation model reflecting the indirect effect of *FALSE POSITIVE* on *PROFESSIONAL SKEPTICISM (PS)* through *MESSAGE SEVERITY*. See the APPENDIX for variable descriptions.

The coefficients (β) are standardized to allow for computation of the indirect effects and we estimate each using:

 $PS (1 \text{ or } 0) = \delta_1 + \beta_1 FALSE POSITIVE + \varepsilon$ $MESSAGE SEVERITY = \delta_2 + \beta_2 FALSE POSITIVE + \varepsilon$ $PS (1 \text{ or } 0) = \delta_3 + \beta_3 FALSE POSITIVE + \beta_4 MESSAGE SEVERITY + \varepsilon$ (3)

We report statistical significance using ***, **, and * which indicate p < 0.01, p < 0.05, and p < 0.10, respectively.

The indirect effect of *MESSAGE SEVERITY* on *PS* is the product of $\beta_2 FALSE POSITIVE*\beta_4 MESSAGE SEVERITY$, and the β coefficient equals -0.53. The confidence interval (CI) is the bias-corrected interval for the estimate of the indirect effect, which we estimated using 5,000 bootstrapped re-samples of the data with replacement. Significance of the indirect effect is indicated if the confidence interval excludes zero.

 β_1 is from the path c and is the *direct effect* of *FALSE POSITIVE* on *PS* (consistent with Table 1) and β_3 is from the path c' and is the *direct effect* of *FALSE POSITIVE* on *PS* controlling for the effect of *MESSAGE SEVERITY*. Overall, Figure 3 shows that participants' perceptions of *MESSAGE SEVERITY* fully mediates the relationship between *FALSE POSITIVE* and *PS*.

FIGURE 4: The Role of RISK-AVERSION



Conditional Effects of *REWARD* on *PS* when *FALSE POSITIVE* rates are higher and *RISK-AVERSION* varies

FALSE	RISK-			95% Confidence
POSITIVE	AVERSION	Effect	p-value	Interval
Higher	18.2462	3.4435	0.009	0.86, 6.02
Higher	38.9528	0.4723	0.498	-0.90, 1.84
Higher	59.6595	-2.4988	0.086	-5.35, 0.35

Figure 4 shows the results a moderated moderation model of the effect of *REWARD* on *PS*, moderated by *FALSE POSITIVE* at different rates of *RISK-AVERSION*. *RISK-AVERSION* values are the mean and +/- 1 standard deviation from the mean. See the APPENDIX for all variable descriptions. The figure demonstrates that the effect of *REWARD* on *PS* is significantly moderated by *FALSE POSITIVE* at different levels dependent upon the auditor's *RISK-AVERSION* level. We used the PROCESS Macro in SPSS, Model 3 (Hayes 2018) to conduct the analyses. We present standardized regression coefficients and the corresponding *p*-values (two-tailed). We report the level of statistical significance using ***, **, and * which indicate p < 0.01, p < 0.05, and p < 0.10, respectively.

TABLE 1: Likelihood of Investigating Difference from Expected Sales (PS)

	Rate of	of False Positives (FALSE POSITIV	<i>E</i>)
-		Lower	Higher	Overall
Consistent Reward for Costly Skepticism (REWARD)		0.29	0.30	0.29
	Absort	(0.46)	(0.47)	(0.46)
	Absent	[28]	[27]	[55]
	REWARD Present	0.69	0.38	0.53
		(0.47)	(0.49)	(0.50)
		[29]	[29]	[58]
		0.49	0.34	
	Overall	(0.50)	(0.48)	
		[57]	[56]	

Panel A: Proportion of Participants Investigating Difference from Expected Sales (Std. Dev) [Cell Size] Rate of False Positives (EALSE POSITIVE)

Panel B: Logistic Regression Results: Tests of H1 and H2

VARIABLES	Professional Skepticism (PS)
CONSTANT	-0.579**
	(2.834)
FALSE POSITIVE (H1)	-0.685**
	(2.907)
REWARD (H2)	1.064***
	(6.947)
Observations	113
Pseudo R-squared	0.113

Panel C: Logistic Regression Results: Test of H3

VARIABLES	Professional Skepticism (PS)
CONSTANT	-0.916***
	(4.798)
FALSE POSITIVE	0.051
	(0.007)
REWARD	1.715***
	(8.749)
FALSE POSITIVE x REWARD (H3)	-1.342**
	(2.729)
Observations	113
Pseudo R-squared	0.143

Panel D: Planned Comparisons: Test of H3

Comparison	df	Contrast	t-statistic	p-value
Effect of REWARD within Lower FALSE POSITIVE	109	-0.40	-3.22	<0.001
Effect of REWARD within Higher FALSE POSITIVE	109	-0.08	-0.66	0.257
Effect of FALSE POSITIVE within REWARD Absent	109	-0.01	-0.08	0.467
Effect of FALSE POSITIVE within REWARD Present	109	0.31	2.50	0.007

Panel A of Table 1 presents descriptive statistics for mean likelihood of exercising skepticism (*PS*) across the four experimental groups. See the APPENDIX for all variable descriptions. In Panels B and C of Table 1, we report the regression coefficients with z-statistics in parentheses. Panel D of Table 1 presents simple effects tests across the four experimental groups. All tests use one-tailed *p-values* and significance is noted at the *** p < 0.01, ** p < 0.05, and * p < 0.1 levels.

TABLE 2: Skeptical Judgments, Documentation, and Skeptical Actions

	FALSE POSITIVE	RECOGNITION	DOCUME	NTATION	PS
DEWADD Alassa	Lower	6.67	0.5	0	0.29
KE WARD Adsent	Higher	6.71	0.1	9	0.30
	Lower	6.14	0.4	5	0.69
KEWARD Present	Higher	6.83	0.1	7	0.38
Panel B: Correlation Mat	rix				
	1	2	3		4
1. FALSE POSITIVE	1				
2. REWARD	0.009	1			
3. RECOGNITION	0.086	-0.043	1		
4. DOCUMENTATION	- 0.314***	-0.037	0.00)9	1
5. <i>PS</i>	- 0.154*	0.247***	0.15	51	0.444***
5. PS Panel C: ANOVA Result	- 0.154* ts (<i>RECOGNITION</i>)	0.247***	0.15	51	0.444***
5. PS Panel C: ANOVA Result VARIABLES	- 0.154* ts (RECOGNITION) SS	0.247***	0.15 MS	51 F	0.444*** <i>p-value</i>
5. PS Panel C: ANOVA Result VARIABLES FALSE POSITIVE	- 0.154* ts (<i>RECOGNITION</i>) SS 3.4	0.247*** 5 df 4 1	0.15 <u>MS</u> 3.44	51 F 0.65	0.444*** <i>p-value</i> 0.211
5. PS Panel C: ANOVA Result VARIABLES FALSE POSITIVE REWARD	- 0.154* ts (<i>RECOGNITION</i>) SS 3.4 1.0	0.247*** 5 <u>df</u> 4 1 7 1	0.15 <u>MS</u> 3.44 1.07	51 F 0.65 0.20	0.444*** p-value 0.211 0.327
5. PS Panel C: ANOVA Result VARIABLES FALSE POSITIVE REWARD FALSE POSITIVE x REV	- 0.154* ts (RECOGNITION) SS 3.4 1.0 WARD 2.6	0.247*** <u>6 df</u> 4 1 7 1 9 1	0.15 <u>MS</u> 3.44 1.07 2.69	<u>F</u> 0.65 0.20 0.51	0.444*** p-value 0.211 0.327 0.238

Panel A: Cell Means for *RECOGNITION*, *DOCUMENTATION*, and *PS*

VARIABLES	DOCUME	INTATION
CONSTANT	-0.02	-0.03
	(-0.07)	(-0.00)
FALSE POSITIVE	-1.42***	-1.48***
	(-3.24)	(-2.38)
REWARD	-0.16	-0.21
	(-0.39)	(-0.39)
FALSE POSITIVE x REWARD		0.12
		(0.14)
Observations	113	113
Pseudo R-squared	0.082	0.082

Panel E: Correlation between *DOCUMENTATION* and *PS*

	$DOCUMENTATION \rightarrow PS$
Lower FALSE POSITIVE— REWARD Absent	.16**
Lower FALSE POSITIVE—REWARD Present	.61***
Higher FALSE POSITIVE— REWARD Absent	.53***
Higher FALSE POSITIVE—REWARD Present	.58***

Table 2 Cont.

Panel A of Table 3 presents descriptive statistics for mean *RECOGNITION*, *DOCUMENTATION* and *PS* across the four experimental groups. Panel B provides a correlation matrix. Panel C presents ANOVA results for *RECOGNITION* and Panel D provides a logistic regression for *DOCUMENT*. Panel E presents the correlations between *DOCUMENTATION* and *PS* across the four experimental groups. See the APPENDIX for all variable descriptions. All tests use two-tailed *p-values* and significance is noted at the *** p < 0.01, ** p < 0.05, and * p < 0.1 levels.