BEHAVIORAL RESEARCH IN ACCOUNTING
Vol. 31, No. 2
Fall 2019
pp. 135–143

American Accounting Association
DOI: 10.2308/bria-52306

The Outcome Effect and Professional Skepticism: A Replication and a Failed Attempt at Mitigation

Joseph F. Brazel
North Carolina State University

Christine Gimbar
DePaul University

Eldar M. Maksymov
Arizona State University

Tammie J. Schaefer
University of Missouri–Kansas City

ABSTRACT: In this research note, we replicate Brazel, Jackson, Schaefer, and Stewart's (2016) study of how auditors evaluate skeptical behavior. Like the original study, we find that evaluators reward audit staff who exercise appropriate levels of skepticism and identify a misstatement (positive outcome). However, when no misstatement is identified (negative outcome), evaluators penalize staff who exercise appropriate levels of skepticism. One factor causing this outcome effect may be that exercising skepticism typically causes budget overages due to additional testing. Hence, we examine whether formally attributing the budget overage to skeptical judgments and actions in the audit budget file reduces outcome effects. However, while replicating the initial effect across three separate studies, we have been unable to reduce this effect. Thus, it is clear that the outcome effect in this context is very robust.

Data Availability: Contact the authors.

Keywords: audit; budget; evaluation; outcome effect; professional skepticism; replication.

I. INTRODUCTION

Despite the important role professional skepticism (PS) plays in relation to audit quality, insufficient auditor skepticism continues to be a global concern (IFAC 2015). Skeptical auditors increase the likelihood that material misstatements in the financial statements are detected, which is important in promoting audit quality, investor confidence, and global financial stability (PCAOB 2012; IFAC 2015). However, exercising skepticism may also come at a cost (e.g., budget overruns, conflicts with management) when additional work is performed to obtain sufficient and appropriate evidence (Nelson 2009; PCAOB 2012). When skepticism incurs a cost but *does not* yield a misstatement, the cost of skepticism may appear, in hindsight, to be unjustified. This can result in supervisors penalizing skepticism when no misstatement is identified—a phenomenon known as the "outcome effect."

Brazel, Jackson, Schaefer, and Stewart (2016) report findings consistent with outcome effects in the evaluations of audit staff. Specifically, when evaluating audit staff, audit seniors penalize (*reward*) skeptical staff who do not detect (*do detect*) a misstatement. In this study, our primary objective is to replicate the outcome effect observed by Brazel et al. (2016) and, as a

We thank Steve Salterio (senior editor), an anonymous reviewer, Scott Fleming, Kathleen Linn, Tim Louwers, Dick Riley, Scott Showalter, representatives from the Center for Audit Quality (CAQ), and workshop participants at the 2016 and 2017 Fall Meetings of the Institute for Fraud Prevention (IFP) for many helpful comments. This research was supported by grants from the CAQ's Access to Audit Personnel program and the IFP. All results, interpretations, and conclusions expressed are those of the authors alone, and do not necessarily represent the views of the CAQ or the IFP. We thank the audit professionals who participated in our experiment.

Editor's note: Accepted by Steven E. Salterio.

Submitted: March 2018
Accepted: August 2018
Published Online: October 2018

secondary objective, to examine whether formally attributing a budget overage to skeptical judgments and actions in the budget file reduces outcome effects in evaluations of audit staff.

Salterio (2014, 1140) discusses how "significant" findings in accounting research should be replicated, lists four criteria for judging the significance of findings, and notes that it often "only takes one of the four to warrant replication."[1] Brazel et al. (2016) satisfy three of the four criteria: the article was originally published in an influential journal, *The Accounting Review,* the study challenges the notion that appropriate PS is rewarded regardless of outcome, and the original article has attracted the attention of regulators and standard setters.[2]

Eighty-nine audit seniors who were experienced in evaluating staff participated in this study. The participants served as evaluators of a skeptical staff person who incurred the costs of skepticism and either did or did not identify a misstatement. We follow the same methods as Brazel et al. (2016). The evidencing of skepticism in the budget file is manipulated between participants by providing an explanation (*not providing an explanation*) for the reason that the staff exercised PS and the resulting increase in testing.

Our replication provides results that are approximately the same as those reported in Brazel et al. (2016). In addition, we replicate Brazel et al.'s (2016) process model, which links outcomes to perceptions to evaluations. We also find that the outcome effect is robust to whether or not the staff's skeptical judgments and actions are explained in the budget file. Finally, our results suggest a benefit of explaining skepticism in the budget when *no misstatement is identified:* evaluators are more likely to perceive audit budget overruns as a "normal cost" of the audit as opposed to "lost time."

The remainder of this paper consists of a brief description of the original study's hypotheses related to the outcome effect, as well as a description of why evidencing skepticism in the budget may mitigate the effect. We then provide our research method and the results of our study. The final section concludes the paper.

II. THEORY AND HYPOTHESES

The Outcome Effect and Professional Skepticism

Professional skepticism is an attitude that includes a questioning mind and a critical assessment of audit evidence (PCAOB 2018). As auditors exercise higher levels of skepticism, they may require more evidence to justify their audit opinions (Nelson 2009). However, skeptical behavior does not always produce the same outcome. For example, an auditor observes a red flag when evaluating audit evidence and exercises an appropriate level of skepticism by performing additional testing. Conducting an investigation would be consistent with exercising appropriate skepticism, but also requires added effort from both the auditor and the client and does not ensure that a misstatement will be found. It is possible that the additional testing leads to an acceptable explanation for the inconsistent evidence observed, such that no audit adjustment is necessary. In sum, the auditor can incur the costs associated with exercising skepticism (e.g., impaired client relations), but may not experience the related "benefit" of identifying a misstatement.

In their review of research on auditor professional skepticism, Hurtt, Brown-Liburd, Earley, and Krishnamoorthy (2013, 56) pose the following empirical question: "Are the audit firms currently evaluating or rewarding skeptical inquiries, regardless of outcomes?'' Brazel et al. (2016) address this question and identify a potential barrier to skepticism: outcome effects that exist in supervisors' *evaluations of skeptical behavior*. Outcome effects refer to situations where the knowledge of an outcome influences evaluators' judgments in the direction of the outcome (Tan and Lipe 1997).

Holding the evidence set and a staff member's skeptical behavior constant, Brazel et al. (2016) find that supervisors (audit seniors) on an engagement team evaluate their staff's skeptical behavior based on whether the staff's investigation of an issue ultimately identifies a misstatement, not whether the behavior itself was appropriate. Supervisors/evaluators reward audit staff who exercise PS that leads to the identification of a misstatement (positive outcome), but penalize staff who exercise PS yet do not identify a misstatement (negative outcome).

One implication of Brazel et al.'s (2016) finding is that the anticipation of outcome effects may, at times, cause auditors to forego skeptical behavior. On the other hand, PS has been a continual focus of standard setters and regulators since Brazel et al.

[1] Salterio (2014, 1140) lists the following four criteria: (1) "where the article was originally published (i.e., on the FT 45 list)," (2) "whether the article challenges other research or accepted beliefs about how the accounting world works," (3) "whether the article attracts the attention of regulators or standard setters," and (4) "whether the article attracts attention from practitioners who try to use the findings 'to make money.'"

[2] For examples, see https://pcaobus.org/News/Events/Documents/11292017-SAG-meeting/professional-skepticism-Tammie-Schaefer.pdf, https://www.ifac.org/global-knowledge-gateway/audit-assurance/discussion/research-insights-auditor-professional, https://twitter.com/iaasb_news/status/674989048600072192, and http://files.iaaer.org/research/IAASB_Executive_Summary_12-3-14_(1).pdf?1421423093

(2016) performed their experiment (e.g., IFAC 2015).[3] Given this focus on skepticism, it is quite possible that the outcome effect observed by Brazel et al. (2016) has been mitigated through improvements in firms' quality controls or training. As such, we replicate the following hypothesis:

H1: Superiors will evaluate skeptical auditors more negatively (*positively*) when they do not (*do*) identify a misstatement.

Mitigating Outcome Effects by Evidencing Professional Skepticism in the Time Budget

Brazel et al. (2016) note that the outcome bias likely exists because, given the costs of skepticism, superiors base their evaluations of PS as though the subordinate auditor should have "known all along" whether a misstatement existed. The study also discusses how "the appropriate criterion for evaluating the auditor is not what they found, but what the evidence suggested they might find" (Brazel et al. 2016, 1581). Superiors exhibiting outcome bias fail to recall that the costs of PS (e.g., budget overruns) were caused by the auditor's decision to *appropriately* exercise professional skepticism, given the evidence set.

Based on Kennedy's (1993) framework for examining how to mitigate biases, we reason that the bias reported by Brazel et al. (2016) is a data-related bias—a bias that occurs when using bad data to make decisions.[4] As Kennedy (1993) explains, data-related biases may stem from a lack of knowledge and/or inadequate memory and could be mitigated by simply "refreshing the judge's memory" and/or endowing the judge with sufficient knowledge. Brazel et al. (2016) tested one such debiaser—consultation with the superior, reasoning that such consultation increases the superior's knowledge of the reasons for skepticism—but found it largely ineffective. The study also reports that a higher level of audit committee support for the engagement team, which should insulate the subordinate auditor from the costs of PS (e.g., budget overruns, strained management relations), does not alleviate the outcome effect. In this study, we examine another potential debiaser that both refreshes memory and provides additional knowledge, as prescribed by Kennedy (1993).

The audit engagement budget file can be used as an effective debiaser (i.e., refreshing the decision maker's memory and providing additional knowledge) by briefly describing the evidence inconsistency, summarizing the additional investigation/testing of the inconsistency, and referencing the audit file where the audit testing is documented. Explaining skeptical behavior in the time budget, regardless of whether a misstatement is identified, directly addresses a chief cost of skepticism (budget overages) and makes the staff's skeptical judgments and acts (the PS *process* as described by Nelson 2009) more salient to the evaluator (vs. simply the outcome). Wu, Shimojo, Wang, and Camerer (2012) observe in a visual exercise that explicitly documenting the *process* by which subordinates make their decisions reduces hindsight bias (similar to outcome effects) in the evaluation process.

Explaining PS that does not yield a misstatement may cause the PS to be viewed more favorably as it has the potential to benefit the audit team by highlighting appropriate applications of PS that might not have otherwise been as explicitly documented. Such documentation may be particularly important in instances where the audit could be inspected by the PCAOB, a regulator concerned about insufficient skepticism being exercised on audits (e.g., PCAOB 2012). In sum, we predict that explaining skepticism in the budget file will mitigate outcome effects in supervisor evaluations of skeptical staff. Stated formally, our hypothesis is as follows:

H2: When audit staff explain their skepticism in the budget file, the outcome effect in the evaluations of skeptical behavior is reduced.

III. METHOD

Purpose

The primary purpose of this experiment is to replicate the outcome effect in auditor evaluations identified by Brazel et al. (2016). The secondary purpose is to determine whether documenting the application of PS in the budget file effectively mitigates the outcome effect.

[3] Brazel et al. (2016) was originally submitted to *The Accounting Review* in December 2013. To illustrate the continued focus on professional skepticism, the topic of professional skepticism was on the agenda of the November 2017 PCAOB Standing Advisory Group Board Meeting (https://pcaobus.org/News/Events/Pages/SAG-meeting-Nov-2017.aspx).

[4] The framework was subsequently tested by Kennedy (1995), and independently tested by Emby and Finley (1997), Libby, Salterio and Webb (2004), and others.

Participants

The experimental participants for this study consisted of 89 audit seniors from two of the eight largest international accounting firms. The experiment was administered while participants attended training sessions. All 89 participants had performed at least one evaluation of a staff auditor. On average, our participants reported having 3.9 years of audit experience, conducted eight performance evaluations of staff auditors under their supervision, and spent 58 percent of their time auditing publicly-traded clients. Forty-nine percent of participants were male.[5]

Brazel et al. (2016) report that their participants consisted of 96 audit seniors from one international accounting firm. Their experiment was also administered while participants attended training sessions. Sixty-nine percent of their participants had performed at least one evaluation of a staff auditor. On average, their participants had 2.83 years of audit experience and had conducted three performance evaluations of staff auditors under their supervision. Non-tabulated t-tests indicate that the participants in the current study have significantly more (p's < 0.01) audit experience, are more likely to have completed at least one evaluation of staff, and have performed a greater number of evaluations than the participants in Brazel et al. (2016).[6]

Description of Experimental Context

The experimental materials were adapted from Brazel et al. (2016). Participants were positioned as the lead senior on the hypothetical audit engagement of Madison, Inc., a publicly-traded manufacturing company with many divisions. Participants were asked to assess the performance of Sam, a staff member under their supervision. Among other tasks, the responsibilities of Sam included performing substantive analytical procedures related to the Sporting Goods division's revenue account (Brazel et al. 2016).[7]

Our experimental context replicates Brazel et al. (2016) with the exception of: (1) keeping the level consultation between the staff and the participant constant, and (2) providing all participants with a budget detailing the staff's budget overage (described below). Brazel et al. (2016) attempt, but are unable, to mitigate the outcome effect through manipulating the level of consultation between the skeptical staff and the evaluator (experimental participant). In our study, we hold the level of consultation constant at the moderate level because discussions with practitioners indicate that is the level most likely to occur in practice. As described by Brazel et al. (2016), the "moderate consultation condition reflects the option to inform their supervisor about a situation and allow the supervisor to provide guidance if needed (this amounts to keeping the supervisor 'in the loop')." Our approach is termed "differentiated replication" and its advantages are discussed by Lindsay and Ehrenberg (1993) and Salterio (2014).

Manipulated Variable—Replication

Consistent with Brazel et al. (2016), the first manipulated variable is whether Sam's investigation uncovered a misstatement (this variable is referred to as *OUTCOME* below). In the no misstatement condition, participants were told the following:

> Sam found that the inconsistency described above was a result of the Sporting Goods division outsourcing some operations overseas. Sam made several inquiries into the matter and collected additional audit evidence, which eventually led to a conclusion that *there were no misstatements in this revenue account*.

In the misstatement condition, participants were told the following:

> Sam found that the inconsistency described above was a result of the Sporting Goods division outsourcing some operations overseas. Sam made several inquiries into the matter and collected additional audit evidence, which eventually led to a conclusion that *a significant misstatement existed in this revenue account* as revenues were being recognized prematurely at the overseas operation.

Manipulated Variable—Extension

The second independent variable manipulated whether or not the application of PS that led to the overage in the revenue account was documented in the budget file (this variable is referred to as *EXPLANATION* below). In the *EXPLANATION* present condition, the following was noted in the budget file:

[5] We obtained institutional review board approval for this study.

[6] While not reported by Brazel et al. (2016), the same percentage of their participants were male (49 percent). Brazel et al. (2016) did not measure the percentage of time their participants spent auditing publicly traded clients.

[7] For additional information about the experimental case study, see Brazel et al. (2016, 1583–1585).

> **Application of professional skepticism:** In performing analytical procedures we identified an inconsistency between (1) the growth in revenue and (2) decreases in the number of employees and production space in the Sporting Goods division (SGD). While SGD's revenue *increased* by 7.15 percent (from $127,235,268 to $136,334,334), the number of SGD's employees and SGD's production space *decreased* by 19.62 percent and 20.27 percent, respectively (from 474 to 381 employees and from 444,000 to 354,000 square feet). The audit guidance pertinent to substantive analytical procedures for revenue includes AS 2110 (which presumes a risk of fraud related to improper revenue recognition) and AS 2305 (which considers unexpected differences in analytical procedures as potentially indicating an increased risk of material misstatement that must be addressed by additional audit procedures). We concluded that this unexpected difference in analytical procedures indicated an increased risk of material misstatement in SGD's revenue account and that additional procedures needed to be performed. The additional procedures included verifying the reliability of NFMs, investigating why the inconsistency existed, and doubling the number of sample selections. **Additional procedures identified a misstatement, and management recorded the resulting audit adjustment (see Adjustment Schedule at workpaper ADJ-80021).**[8]

In the no *EXPLANATION* condition, there was no explanation in the budget file documenting why there was an overage for the revenue testing.[9]

Dependent Variables

Similar to Brazel et al. (2016), we measured participants' evaluations of Sam's performance and the components of the Lipe (1993) outcome effect model. The primary dependent variable is the participant's overall performance evaluation (*EVAL*), which is their response to the question "How would you evaluate Sam's overall performance?" Responses were provided on an 11-point scale ranging from −5 to +5 with the left endpoint labeled "below expectations," the right endpoint labeled "above expectations," and the midpoint labeled "met expectations."

IV. RESULTS

Panel A of Table 1 reports means for the dependent variable in our experimental conditions, and Figure 1 graphs those means. Consistent with H1 and the results of Brazel et al. (2016), in Table 1 *OUTCOME* appears to have a large influence on performance evaluations. Indeed, the means for "No Misstatement" and "Misstatement" in Panel A (1.65 and 3.26, respectively) are approximately the same as those observed by Brazel et al. (2016) in their moderate consultation condition (1.69 and 3.18, respectively). However, contrary to H2, Panel A of Table 1 and Figure 1 indicate that explaining PS in the budget does not significantly reduce the outcome effect.

Panel B of Table 1 provides the analysis of variance (ANOVA) results to formally test H1 and H2. Replicating Brazel et al. (2016) and supporting H1, the ANOVA results reveal that the overall performance evaluation is strongly influenced by *OUTCOME* (F-statistic = 20.28, p-value < 0.001). *EXPLANATION,* however, does not seem to significantly influence the overall performance evaluation, either directly (F-statistic = 0.09, p-value = 0.771) or via an interaction with *OUTCOME* (F-statistic = 0.04, p-value = 0.847). Thus, the outcome effect is not mitigated by evidencing PS in the budget.

[8] Note that the last sentence in bold was only present in the *misstatement* condition.

[9] Eighty-six (96.6 percent) of the 89 participants correctly completed the manipulation check question for the *OUTCOME* manipulation, and 72 (80.9 percent) of the 89 participants correctly completed the manipulation check question for the *EXPLANATION* manipulation. Analyzing only participants that correctly completed both manipulation check questions does not alter the conclusions drawn. Thirteen (76 percent) of the 17 *EXPLANATION* manipulation failures occurred in the no *EXPLANATION* conditions, such that participants who were *not* shown an explanation in the budget file reported that they *had* viewed a detailed explanation of the budget overage. This may be a result of the following features of the instrument: When designing the instrument, we were careful to avoid any mention of a "detailed explanation of budget overages" prior to the collection of our primary dependent variables so as not to bias responses from those in the no *EXPLANATION* condition. As such, the placement of our manipulation check questions in the middle of the 11 pages of post-experimental questions may have resulted in a lack of attention. Further, two elements of our no *EXPLANATION* condition may have led participants to believe that they *had* reviewed a more "detailed" budget than they are accustomed to seeing in practice. First, our no *EXPLANATION* condition does provide a brief explanation just above the budget itself that says, "Sam's additional procedures in the Sporting Goods division caused Sam to go significantly over budget and resulted in the overall Madison revenue hours being over-budget, reflected below." This explanation may have provided the "detail" that participants reported in the manipulation check. Second, while sparse, our no *EXPLANATION* condition does provide budgeted hours by line item (e.g., Cash, Accounts Receivable, etc.), which could be more detailed than what the participants had experienced in practice, thus resulting in the manipulation check failure. However, we constructed this budget from a budget obtained from an actual audit client and had several audit managers, partners, and a former partner review the budget for realism.

TABLE 1

Descriptive Statistics and Testing of Hypotheses

Panel A: Performance Evaluation in Experimental Conditions

Misstatement Conditions (*OUTCOME*)	Explaining Skepticism Conditions (*EXPLANATION*)		
	No Explanation	Explanation	Marginal Means
No Misstatement			
Mean	1.63	1.67	1.65
Std. Dev.	1.89	1.90	1.88
n	19	24	43
Misstatement			
Mean	3.17	3.35	3.26
Std. Dev.	1.70	1.15	1.44
n	23	23	46
Marginal Means			
Mean	2.48	2.49	2.48
Std. Dev.	1.93	1.78	1.84
n	42	47	89

Panel B: ANOVA Results for Performance Evaluation

Source of Variation	DF	SS	F-statistic	p-value
OUTCOME	1	57.33	20.284	< 0.001
EXPLANATION (*EXP*)	1	0.24	0.09	0.771
OUTCOME $*$ *EXP*	1	0.11	0.04	0.847
Error	85	2.83		

R^2 (%) = 19.4
Model F-statistic = 6.83 ($p < 0.001$)

The dependent variable is participants' performance evaluation, which is their response to the question "How would you evaluate Sam's overall performance?" (responses are provided on an 11-point scale ranging from −5 to +5 with the left endpoint labeled "below expectations," the right endpoint labeled "above expectations," and the midpoint labeled "met expectations"). The manipulated variables are defined as follows: *OUTCOME* is the outcome of Sam's investigation, which is manipulated between participants as either (1) Sam found that there was no misstatement (coded as 1) or (2) Sam found that there was a significant misstatement (coded as 0); *EXPLANATION* is whether or not the application of professional skepticism that led to the overage in the revenue account was documented in the budget file.

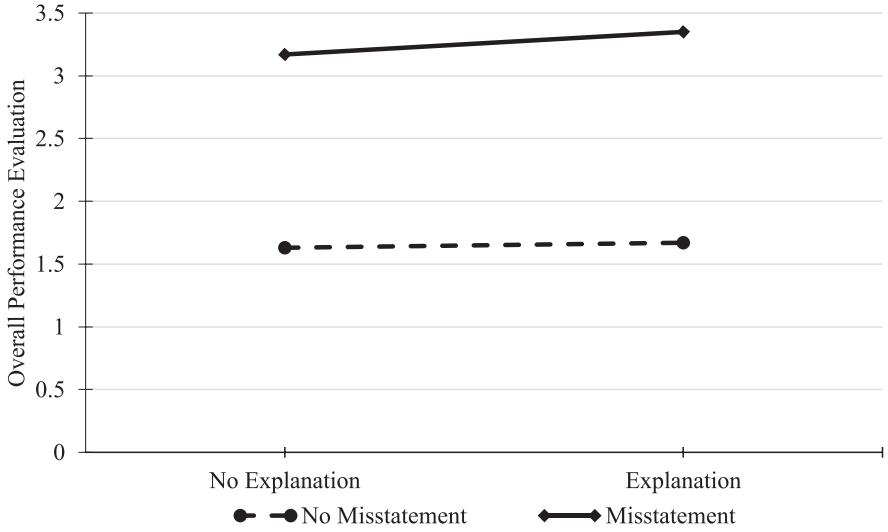
Additional Analyses

We also replicate the process model proposed by Lipe (1993) and tested in Brazel et al. (2016). The model predicts that the outcome of Sam's investigation (*OUTCOME*) will affect the evaluator's perception of whether the audit team got some benefit from Sam's time spent investigating (*BENEFIT*). *BENEFIT* then influences whether the evaluator frames the cost of the investigation as "lost time" or a "normal cost" of the audit (*FRAME*). In turn, the decision frame adopted by the evaluator affects the overall performance evaluation of Sam (*EVAL*).[10]

We replicate Brazel et al. (2016) as shown in Figure 2. When no misstatement is found, evaluators perceive little benefit from the time Sam spent investigating and view it as "lost time." In turn, the loss frame contributes to a lower evaluation of skepticism that does not ultimately identify a misstatement. In Figure 2 we do observe one difference between our model results and those reported by Brazel et al. (2016)—namely, a significant negative effect for *OUTCOME* on *FRAME*. Consistent with

[10] *EVAL* is the participant's performance evaluation and *OUTCOME* is manipulated as described in the Method section. *BENEFIT* is measured as participants' responses to the question "Do you feel that the audit team got some benefit from the time that Sam spent to investigate the inconsistency between the growth in revenues and the nonfinancial measures?" Responses are provided on an 11-point scale with the left endpoint labeled "there was no benefit" and the right endpoint labeled "there was a benefit." *FRAME* is measured as participants' responses to the question "Do you view the time that Sam spent investigating the inconsistency between the growth in revenues and the non-financial measures as "lost time" or a "normal cost" of an audit?" Responses are provided on an 11-point scale with the left endpoint labeled "lost time" and the right endpoint labeled "normal cost."

FIGURE 1
Graph of Cell Means for Performance Evaluation in Experimental Condition



The dependent variable is participants' overall performance evaluation, which is their response to the question "How would you evaluate Sam's overall performance?" (responses are provided on an 11-point scale ranging from −5 to +5 with the left endpoint labeled "below expectations," the right endpoint labeled "above expectations," and the midpoint labeled "met expectations"). The manipulated variables are defined as follows: *OUTCOME* is manipulated between participants as either (1) Sam found that there was no misstatement or (2) Sam found that there was a significant misstatement, and *EXPLANATION* is whether or not the application of professional skepticism that led to the budget overage in the revenue account was documented in the budget file.

what one would expect, when the outcome of PS is not the identification of a misstatement, evaluators are more likely to frame the time spent as "lost time."

With respect to the role of *EXPLANATION* and evaluator perceptions, in a non-tabulated ANOVA we observe that *EXPLANATION* moderates the effect of *OUTCOME* on *FRAME* (p-value = 0.058, one-tailed). Specifically, participants in the "No Misstatement" condition who saw PS explained in the budget file (versus no explanation) were more likely to view the budget overage as a "normal cost" of the audit (p-value = 0.058, one-tailed). *EXPLANATION* had no effect on *OUTCOME* in the "Misstatement" condition (p-value = 0.668, two-tailed). Thus, while *EXPLANATION* did not mitigate the outcome effect, it does have the potential to change how evaluators frame the costs of PS.

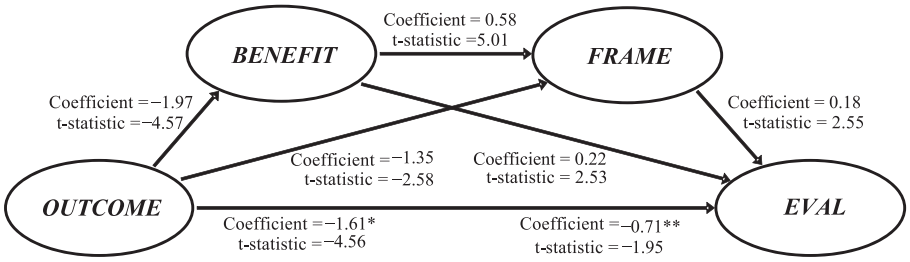
V. DISCUSSION AND CONCLUSIONS

The findings of Brazel et al. (2016) and this study confirm that outcome effects play a substantial role when evaluating professional skepticism. Brazel et al. (2016) observe that both moderate and extensive consultation between the subordinate and the evaluator does not mitigate the outcome effect, nor does more extensive support from the audit committee. We find that that the outcome effect has persisted despite the focus on PS by firms, regulators, and standard setting (e.g., PCAOB 2012; IFAC 2015).[11] We also replicate the outcome effect observed by Brazel et al. (2016) with audit seniors with significantly more experience performing evaluations of staff. As such, our results suggest that outcome effects in evaluations of PS may not be mitigated by more experience evaluating subordinates. Determining whether such outcome effects persist to the levels of audit manager or partner would be a fruitful area for further research.

Though explaining PS can cause evaluators to perceive budget overages as "normal" rather than "lost time" when no misstatement is identified, it does not appear sufficient to mitigate the outcome effect on the evaluation of the subordinate. While the intent of our explanation condition was to make the PS process salient to evaluators, it is also possible that our manipulation enhanced the *justification* of PS as our manipulation cited professional standards to support the staff's application of PS. While one would expect that both saliency and justification of the PS process would reduce outcome effects, it is possible that the inclusion of both in our experimental manipulations may have interacted and in some way that biased against

[11] Note also that the current study was supported by grants from the Center for Audit Quality's Access to Audit Personnel program and the Institute for Fraud Prevention.

FIGURE 2
Three-Path Mediation Model for Auditors' Performance Evaluations (n = 89)



* Total effect (relationship between *OUTCOME* and *EVAL* without indirect paths included in model).

** Direct effect (relationship between *OUTCOME* and *EVAL* with indirect paths included in model).

OUTCOME is the independent variable, which is manipulated between participants as either Sam found that there was no misstatement (coded as 1) or Sam found that there was a significant misstatement (coded as 0); *BENEFIT* is measured as participants' response to the question "Do you feel that the audit team got some benefit from the time that Sam spent to investigate the inconsistency between the growth in revenues and the nonfinancial measures?" (responses are provided on an 11-point scale with the left endpoint labeled "there was no benefit" and the right endpoint labeled "there was a benefit"); *FRAME* is measured as participants' response to the question "Do you view the time that Sam spent investigating the inconsistency between the growth in revenues and the non-financial measures as 'lost time' or a 'normal cost' of an audit?" (responses are provided on an 11-point scale with the left endpoint labeled "lost time" and the right endpoint labeled "normal cost"); and *EVAL* is the dependent variable, which is measured as participants' response to the question "How would you evaluate Sam's overall performance?" (responses are provided on an 11-point scale with the left endpoint labeled "below expectations," the right endpoint labeled "above expectations," and the midpoint labeled "met expectations").

finding a result consistent with H2. In addition, as noted in footnote 9, 19 percent of participants failed the manipulation check associated with the explanation manipulation. As such, our manipulation may have caused some confusion amongst a portion of our sample that biased against our finding mitigation of the outcome effect.

It is also possible that explanations of PS would be more effective if more than one incidence of skepticism without an identified misstatement were provided in the budget file. Such documentation would highlight that skepticism without a resulting audit adjustment is common. Another possibility is that the bias reported by Brazel et al. (2016) could be largely effort-related, contrary to our assumption that the bias is data-related. If this bias is effort-related, according to the framework suggested by Kennedy (1993), it may be mitigated by increased supervisor accountability. Future research could examine these possibilities.

Given that the outcome effect *vis-à-vis* the evaluation of PS appears to be robust and pervasive in the audit setting, future research can also examine how this effect is influencing how skepticism is applied by staff in the field. For example, after being penalized multiple times for exhibiting skepticism that ultimately does not identify a misstatement, are audit staff more apt to ignore or not search for evidence inconsistencies? In addition, given that *some* audit supervisors are less likely to exhibit outcome effects in their evaluations, future research can identify the traits and experiences that are associated with auditors that reward appropriate skepticism, *regardless of the outcome*. Furthermore, audit staff work for multiple supervisors on multiple engagements throughout the year. How do these staff react when experiencing variation in the extent to which their supervisors reward and penalize appropriate skepticism based on its outcome? Last, our study does not investigate conditions where the auditor decides to *not* exercise skepticism (e.g., identifying a red flag, yet not investigating). Being aware of, but not appropriately considering, contradictory audit evidence is a behavior consistent with PCAOB inspection findings (PCAOB 2012; PCAOB 2017a; PCAOB 2017b). We believe that studying the decision to not apply skepticism is an intriguing path for future research.

REFERENCES

Brazel, J. F., S. B. Jackson, T. J. Schaefer, and B. W. Stewart. 2016. The outcome effect and professional skepticism. *The Accounting Review* 91 (6): 1577–1599. https://doi.org/10.2308/accr-51448

Emby, C., and D. Finley. 1997. Debiasing framing effects in auditors' internal control judgments and testing decisions. *Contemporary Accounting Research* 14 (2): 55–77. https://doi.org/10.1111/j.1911-3846.1997.tb00527.x

Hurtt, R. K., H. Brown-Liburd, C. E. Earley, and G. Krishnamoorthy. 2013. Research on auditor professional skepticism: Literature synthesis and opportunities for future research. *Auditing: A Journal of Practice & Theory* 32 (Supplement): 45–97. https://doi.org/10.2308/ajpt-50361

International Federation of Accountants (IFAC). 2015. *Invitation to Comment: Enhancing Audit Quality in the Public Interest: A Focus on Professional Skepticism, Quality Control and Group Audits*. New York, NY: IFAC.

Kennedy, J. 1993. Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research* 31 (2): 231–245. https://doi.org/10.2307/2491272

Kennedy, J. 1995. Debiasing the curse of knowledge in audit judgment. *The Accounting Review* 70 (2): 249–273.

Libby, T., S. E. Salterio, and A. Webb. 2004. The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review* 79 (4): 1075–1094. https://doi.org/10.2308/accr.2004.79.4.1075

Lindsay, R. M., and A. S. C. Ehrenberg. 1993. The design of replicated studies. *The American Statistician* 47 (3): 217–228.

Lipe, M. G. 1993. Analyzing the variance investigation decision: The effects of outcomes, mental accounting, and framing. *The Accounting Review* 68 (4): 748–764.

Nelson, M. 2009. A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice & Theory* 28 (2): 1–34. https://doi.org/10.2308/aud.2009.28.2.1

Public Company Accounting Oversight Board (PCAOB). 2012. *Maintaining and Applying Professional Skepticism in Audits. Staff Audit Practice Alert No. 10 (SAPA 10)*. Washington, DC: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2017a *Standing Advisory Board Meeting: Panel Discussion—Professional Skepticism*. Washington, DC: PCAOB. Available at: https://pcaobus.org/News/Events/Documents/11292017-SAG-meeting/Professional-Skepticism-Briefing-Paper.pdf

Public Company Accounting Oversight Board (PCAOB). 2017b. *Staff Inspection Brief: Information about 2017 Inspections*. Washington, DC: PCAOB. Available at: https://pcaobus.org/Inspections/Documents/inspection-brief-2017-3-issuer-scope.pdf

Public Company Accounting Oversight Board (PCAOB). 2018. *Due Professional Care in the Performance of Work. AS 1015*. Washington, D.C: PCAOB.

Salterio, S. E. 2014. We don't replicate accounting research—Or do we? *Contemporary Accounting Research* 31 (4): 1134–1142. https://doi.org/10.1111/1911-3846.12102

Tan, H.-T., and M. G. Lipe. 1997. Outcome effects: The impact of decision process and outcome controllability. *Journal of Behavioral Decision Making* 10 (4): 315–325. https://doi.org/10.1002/(SICI)1099-0771(199712)10:4<315:AID-BDM257>3.0.CO;2-L

Wu, D.-A., S. Shimojo, S. W. Wang, and C. F. Camerer. 2012. Shared visual attention reduces hindsight bias. *Psychological Science* 23 (12): 1524–1533. https://doi.org/10.1177/0956797612447817