WORKING PAPER

An Unintended Consequence of Full Population Testing on Auditors' Professional Skepticism

Xiaoxing Li
Joseph F. Brazel
Anna Gold

KEY TAKE-AWAYS

The emergence of data analytics allows auditors to test entire populations of data, rather than relying solely on sampling methods. While full population testing increases the sufficiency¾or quantity¾of evidence examined, it does not necessarily eliminate its lack of appropriateness¾or quality. In particular, full population testing typically relies on client-internal data, which are vulnerable to management manipulation, potentially reducing their appropriateness. Therefore, auditors must remain skeptical when subsequent, more appropriate evidence from external sources contradicts a client's financial reporting. We examine whether auditors employing full population testing mistakenly substitute their assessment of evidence sufficiency for their evaluation of evidence appropriateness, leading them to view client-internal evidence as more appropriate than auditors using sample testing.

Consequently, auditors using full population testing may be less likely to act skeptically when subsequent, more appropriate external evidence reveals a fraud red flag. In an experiment, we find that auditors using full population testing, compared to sample testing, are less likely to exercise skeptical actions when a subsequent external industry growth trend reveals a fraud red flag. We also posit that this unintended consequence is exacerbated when full population testing results are visualized (versus tabulated). However, our findings do not support this prediction.

2024/10 – 30

Project Number:
2021B01

F/AR Foundation *for* Auditing Research

An Unintended Consequence of Full Population Testing on Auditors' Professional Skepticism

XIAOXING LI
Department of Accounting
School of Business and Economics
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
x4.li@vu.nl

JOSEPH F. BRAZEL
Department of Accounting
Poole College of Management
North Carolina State University
Raleigh, North Carolina, USA
jfbrazel@ncsu.edu

ANNA GOLD*
Department of Accounting
School of Business and Economics
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
anna.gold@vu.nl

October 2024

We appreciate the helpful comments provided by Tina Carpenter, Marcus Doxey, Kris Hardies, Rick Hatfield, Justin Leiby, Kathleen Linn, Nikki MacKenzie, Jenny McCallen, Mark Peecher, Qinnan Ruan, Steve Salterio, Jaime Schmidt, Quinn Swanquist, Mike Wilkins, and participants at a 2022 Foundation for Auditing Research (FAR) PhD Seminar, the 2023 European Audit Research Network (EARNet) PhD Workshop, a 2023 Vrije Universiteit Amsterdam School of Business and Economics Experimental Seminar, the 2024 University of Alabama Accounting Research Conference, the 2024 European Network for Experimental Accounting Research (ENEAR) Conference, a 2024 Vrije Universiteit Amsterdam Accounting Department Brownbag Seminar, the 10th Workshop on Audit Quality, a 2024 FAR Masterclass, and a 2024 NHH Norwegian School of Economics Seminar. This study was supported by a research grant from the FAR. All results, interpretations, and conclusions expressed are those of the authors alone, and do not represent the views of the FAR. We thank the audit professionals who contributed to our study. We are also grateful for the research assistance provided by Blake Fasciolo, Carley Hazelton, Zach Helms, Parker Holmes, Jude Lomax, Claryssa Tarigan, and Muhser Thayousoe.

* Corresponding author

An Unintended Consequence of Full Population Testing on Auditors' Professional Skepticism

ABSTRACT

The emergence of data analytics allows auditors to test entire populations of data, rather than relying solely on sampling methods. While full population testing increases the *sufficiency*—or quantity—of evidence examined, it does not necessarily eliminate its lack of *appropriateness*—or quality. In particular, full population testing typically relies on client-internal data, which are vulnerable to management manipulation, potentially reducing their appropriateness. Therefore, auditors must remain skeptical when subsequent, more *appropriate* evidence from external sources contradicts a client's financial reporting. We examine whether auditors employing full population testing mistakenly substitute their assessment of evidence *sufficiency* for their evaluation of evidence *appropriateness*, leading them to view client-internal evidence as more appropriate than auditors using sample testing. Consequently, auditors using full population testing may be less likely to act skeptically when subsequent, more appropriate external evidence reveals a fraud red flag. In an experiment, we find that auditors using full population testing, compared to sample testing, are less likely to exercise skeptical actions when a subsequent external industry growth trend reveals a fraud red flag. We also posit that this unintended consequence is exacerbated when full population testing results are visualized (versus tabulated). However, our findings do not support this prediction.

JEL codes: M40, M41, M42, M49

Keywords: attribute substitution, external evidence, fraud, full population testing, professional skepticism

1. *Introduction*

Auditing standards require auditors to obtain sufficient appropriate audit evidence as the basis for their audit opinion. *Sufficiency* relates to the *quantity* of evidence (e.g., sample size), while *appropriateness* pertains to its *quality* (e.g., evidence obtained from independent sources) (e.g., Proposed ISA 500 Revised, IAASB [2022a], AS 1105, PCAOB [2024a]). Emerging technologies, such as data analytics, have expanded auditors' evidence by shifting audits from sampling methods to testing entire populations or every transaction recorded in an account (e.g., KPMG [2015], Deloitte [2016], BDO [2022], PwC [2022], EY [2023], PCAOB [2024b, c]), thereby increasing the sufficiency of the evidence examined. Full population testing (FPT) indeed offers numerous advantages over sampling methods, including identifying *all* potential errors in a population, enhancing efficiency, and ultimately improving audit quality (e.g., Hoogduin et al. [2015], Johnson and Wiley [2022], Huang et al. [2022]). For example, during substantive testing, auditors can leverage analytics to perform automated three-way matches on the entire population of a client's sales balance, verifying sales prices and quantities by examining sales orders, invoices, and shipping documents (e.g., AICPA [2017]).

However, testing the entire population of a given source of evidence rather than a sample does not necessarily eliminate its lack of appropriateness. Indeed, "obtaining more of the same type of audit evidence may not compensate for its lack of appropriateness" (AU-C Section 500, AICPA [2021a, p. 435]).[1] In particular, FPT relies heavily on electronic data sourced exclusively from clients' internal information systems (e.g., Freiman, Kim, and Vasarhelyi [2022], Huang et al. [2022], PCAOB [2023]), such as client invoices. Hence,

[1] Consistent articulation can be found in Proposed ISA 500 Revised (IAASB [2022a]) and AS 1105 (PCAOB [2024a]). Proposed ISA 500 Revised notes that "obtaining more audit evidence, however, may not compensate for its poor quality" (IAASB [2022a, p. 29] and AS 1105 states that "obtaining more of the same type of audit evidence, however, cannot compensate for the poor quality of that evidence" (PCAOB [2024a, para. 05]).

compared to sample testing, FPT primarily enables auditors to obtain more sufficient evidence generated from client-internal sources (*internal* evidence). FPT is less likely to be applied to external evidence in practice.[2] Importantly, internal evidence is more vulnerable to management manipulation, potentially reducing its appropriateness (e.g., PCAOB [2024a]). Thus, despite testing a larger volume of internal evidence through FPT, relying heavily on internal evidence is particularly problematic when external evidence subsequently contradicts a client's financial reporting. For example, a client's unusually rapid sales growth, relative to industry peers, could signal a fraud red flag that warrants further investigation (e.g., IAASB [2022a], Brazel, Jones, and Lian [2023]). Heavily relying on internal evidence may, therefore, impair fraud detection (e.g., IAASB [2022a, b]). As such, even when testing an entire population of internal evidence, adequate consideration of more appropriate external evidence remains particularly critical for professional skepticism and fraud detection (e.g., IAASB [2022a]).

In this study, we posit that the use of FPT on internal evidence may trigger an *attribute substitution bias* in auditors' evaluations of evidence (e.g., Kahneman and Frederick [2002]), reducing their skeptical actions when subsequent red flags from more *appropriate* sources (e.g., external evidence) are present. Specifically, we predict that auditors using FPT, compared to those using sample-based testing, are more likely to substitute their assessment of evidence *sufficiency* (i.e., quantity) for their evaluation of evidence *appropriateness* (i.e., quality). The possibility of attribute substitution bias arises from a simplification in evaluating evidence sufficiency (i.e., quantity) for a given audit procedure when performing

[2] We posit that the inclusion of *external* evidence in FPT, while theoretically possible, is highly unlikely in audit practice to date. Discussions with an audit director specializing in digital transformation at an international non-Big Four audit firm revealed that FPT, like the three-way matches we employ in our experiment, currently exclusively relies on standardized data drawn from the *client's* systems. Integrating data directly from external sources (e.g., bank statements) would present considerable challenges, largely due to the lack of standardization. Also, such tests would yield a substantial number of exceptions that, after further manual investigation, would ultimately be deemed false positives due to data standardization issues (e.g., a customer payment on multiple invoices).

FPT. In other words, by testing 100% of the client's transactions, FPT allows auditors to easily conclude that they have obtained sufficient evidence for a given procedure, making evidence sufficiency a highly objective and accessible attribute. In contrast, the *appropriateness* of evidence, such as its reliability or relevance, remains a subjective, less accessible attribute, regardless of whether the evidence comes from testing the entire population or a sample (e.g., Bamber [1983], Rebele, Heintz, and Briden [1988], Knechel and Messier [1990], Hirst [1994], PCAOB [2023]). As a result, we predict that auditors using FPT will substitute their assessment of sufficiency for appropriateness, leading them to an inflated assessment of internal evidence appropriateness compared to auditors employing traditional sample testing. This inflated perception of appropriateness, in turn, is likely to reduce their skepticism when a fraud red flag, such as an inconsistent industry trend, is subsequently exposed by more appropriate *external* evidence.

We also examine if presenting FPT results in a visualized versus a tabulated format further exacerbates the negative effect of FPT on auditors' skeptical actions. Visualizations, such as graphs or charts, are increasingly used in practice to facilitate the communication and comprehension of data analytic test results (e.g., AICPA [2017], Austin et al. [2021], Henry et al. [2023]). Visualizations can benefit auditor judgments by, for example, facilitating the identification of evidence inconsistencies (e.g., Higginbotham, Nash, and Demeré [2021], Baaske, Eulerich, and Wood [2023]). However, there are also potential drawbacks associated with auditors' use of visualizations (e.g., Baaske [2021], Chang and Luo [2021]). Visual representations of FPT results, compared to more traditional tabulated formats, may further increase the accessibility of evidence sufficiency and hence ease its assessment under FPT (e.g., Alonso et al. [1998]). Visualized formats may also cause faster information processing, further reducing the likelihood of the conscious processing that could challenge the substitution bias (e.g., Benbasat and Dexter [1986], Townsend and Kahn [2014], Backof,

Carpenter, and Thayer [2018], Eberhard [2023]). Therefore, we predict that presenting FPT results in a visualized versus tabulated format will exacerbate the negative effect of FPT on auditors' skeptical actions when external evidence subsequently reveals a fraud red flag.

We employed a between-participants experiment to test our predictions. Participants were practicing audit professionals tasked with a year-end substantive analytical procedure for a sales account. The hypothetical audit case was adapted from Brazel, Leiby, and Schaefer [2022] and Barr-Pulliam et al. [2023]. In the introduction, participants were told that they had *previously* performed two three-way matches of the client's sales transactions (i.e., *internal* evidence) during interim testing. Participants in the *FPT* condition were informed that these three-way matches covered the entire population of sales transactions from 01/01/2022 to 10/31/2022, while those in the *Sample Testing* condition learned that the matches were based on a sample of the sales transactions from the same time span. All participants were then provided with identical final results from the interim tests (i.e., an immaterial adjustment of the sales account).[3] Additionally, we manipulated whether participants received the interim testing results in a visualized or tabulated format.

Participants were then asked to perform the year-end substantive analytical procedure for the sales account. We embedded an *external* evidence inconsistency in a dataset different from the one used to derive the interim three-way matches. Participants learned that this dataset was used specifically for performing the year-end substantive analytical procedure. In this dataset, the client's current-year sales growth exceeded the industry growth rate by approximately 17 percent, reflecting a common red flag for fraud firms observed by Brazel, Jones, and Lian [2023]. All other data relevant to the year-end substantive analytical procedure, such as prior year balances and ratios, were consistent with the client's reported

[3] Importantly, participants in the *Sample Testing* condition were informed that they extrapolated the sample testing results to the entire population, including an adjustment for sampling risk.

sales balance. Participants developed an expectation for the sales account and determined whether additional work related to the analytical procedure would be necessary. We measured skeptical actions based on whether participants would inquire of client management and/or inform their audit manager about the industry growth red flag.

We find that auditors using FPT, compared to those using sample testing, are less inclined to inquire of client management or inform their manager about the industry growth red flag. Specifically, 48 percent of the auditors in *FPT* condition chose to act skeptically, whereas 65 percent of the auditors in the *Sample Testing* condition decided to follow up on the fraud red flag. The logistic regression coefficient indicates that auditors in the *FPT* condition are 52 percent less likely to address the industry growth red flag compared to auditors in the *Sample Testing* condition. In supplemental analyses, we illustrate the attribute substitution bias at work by providing evidence that FPT inflates auditors' assessments of internal evidence appropriateness because they perceive that FPT increases sufficiency. Finally, although the overall interaction pattern is consistent with our expectation, we find no statistical evidence that presenting the FPT results in a visualized format exacerbates the negative effect of FPT on auditors' skeptical actions. This is an encouraging finding, given that visualizing FPT results is the format typically used in practice.

This study contributes to the growing literature on audit technologies and their potential effects on auditors' professional skepticism (e.g., Rose et al. [2017], Anderson, Hobson, and Peecher [2020], Baaske, Eulerich, and Wood [2023]). Given the continued importance of auditors exercising professional skepticism, especially in higher fraud risk settings (e.g., Hobson et al. [2017], McAllister, Blay, and Kadous [2021]), our study offers important practical implications and theoretical insights into an unintended effect of FPT on auditors' evaluations of audit evidence and their application of professional skepticism.

Our study also contributes to the understanding of auditors' use of external evidence, which has gained increasing recognition in both the academic literature and among regulators and standard setters (e.g., PCAOB [2021], Brazel, Jones, and Lian [2023], Proposed ISA 240 Revised, IAASB [2024]). Moreover, technological advancements like artificial intelligence are facilitating growing access to information from various external sources (e.g., IAASB [2022b]). However, concerns about auditors failing to obtain sufficient appropriate audit evidence through technology-assisted audit procedures have prompted recent amendments to standards (PCAOB [2024b, c]). Our findings are informative as they suggest that despite these advancements, auditors may not fully capitalize on the expanded availability of external evidence to potentially improve audit quality when employing FPT to obtain and examine internal evidence. Thus, this study informs standard setters and regulators about how auditors evaluate internal evidence and utilize external evidence in the context of advanced audit technologies (e.g., PCAOB [2021]). Audit firms' quality control systems and training may need to address this issue to ensure that advancements in data analytics not only enhance audit efficiency, but also promote the reliance on more appropriate evidence and ultimately higher-quality audits.

More broadly, the current study makes a significant contribution to the literature on auditors' evaluations of audit evidence. Failing to obtain sufficient appropriate audit evidence is among the leading causes of audit deficiencies (e.g., Beasley, Carcello, and Hermanson [2001], Church and Shefchik [2012], Beasley et al. [2013], Rajgopal, Srinivasan, and Zheng [2021]). Although auditing standards clearly define and distinguish between the two fundamental underpinning constructs for audit evidence evaluation, considerable ambiguity remains regarding how auditors assess the sufficiency and the appropriateness of evidence in practice (e.g., Glover, Taylor, and Wu [2019], Altiero, Baudot and Hazgui [2023]). Auditing standards establish a *unidirectional* relationship between evidence appropriateness and

evidence sufficiency. Specifically, "the quantity of audit evidence needed is affected by its quality" and "as the quality of the evidence increases, the need for additional corroborating evidence decreases" (PCAOB [2023, p. 56]). "Obtaining more of the same type of audit evidence, however, cannot compensate for the poor quality of that evidence" (PCAOB [2024a, para. 06]). By applying attribute substitution theory, we explain why auditors may mistakenly evaluate more sufficient but lower-quality evidence as more appropriate. Our findings therefore suggest that auditors may misinterpret the relationship between evidence appropriateness and evidence sufficiency as *bidirectional*. In other words, while auditing standards only state that evidence appropriateness affects the level of sufficiency needed, our participants' perceptions of evidence sufficiency impacted their beliefs about the appropriateness of the evidence they examined. This finding also helps explain why auditors often respond ineffectively to heightened fraud risk by merely increasing the *extent* of the same audit procedure (e.g., increasing the sample size to increase the sufficiency of evidence), rather than making modifications to the audit procedure or the *nature* of testing to obtain more appropriate evidence (e.g., Hammersley [2011], Hammersley, Johnstone, and Kadous [2011]).

2. Theory and Hypothesis Development

2.1 FULL POPULATION TESTING AND AUDIT EVIDENCE

Data analytics "enable more extensive testing of electronic transactions and account files. Such techniques can be used […] to test an entire population instead of a sample" (IAASB [2022a], p. 60]). This quote highlights an ongoing shift in audit methodology, reflecting a growing trend toward testing entire populations of account balances, which reduces dependence on sampling methods (e.g., KPMG [2015], PwC [2022], EY [2023], PCAOB [2024b, c]). By combating the limitations of sampling (e.g., Wurst, Neter, and Godfrey [1991], Hoogduin et al. [2015], Huang et al. [2022]), the adoption of full population testing

(FPT) is expected to improve audit quality (e.g., AICPA [2017], Barr-Pulliam [2019], FRC [2020], Amato [2021], Barr-Pulliam, Brown-Liburd, and Sanderson [2021], Tysiac [2022]). Additionally, the transition to FPT offers efficiency gains over manual testing methods, especially for data in electronic formats where audit procedures are repetitive and amenable to automation (e.g., Johnson and Wiley [2022], IAASB [2022a]). For example, during substantive testing, auditors can leverage analytics to perform automated three-way matches across the entire population of a client's sales transaction data available in electronic form, verifying sales quantities or prices with the examination of sales orders, sales invoices, shipping documents, and price lists (e.g., AICPA [2017], ISA 315 Revised, IAASB [2019]).

FPT currently relies heavily on electronic data extracted from the client's internal information systems, such as client invoices and shipping documents (e.g., Freiman, Kim, and Vasarhelyi [2022], Huang et al. [2022]). This reliance introduces a potential drawback: "a greater susceptibility to management bias may exist when information is generated from internal sources" (AICPA [2021a, p. 440]). Therefore, although FPT substantially increases the *quantity* of evidence, it does not necessarily improve its *quality*, as the evidence is still primarily generated internally from the client's information systems. Auditing standards reflect this concern, noting that "obtaining more of the same type of audit evidence may not compensate for its lack of appropriateness" (AICPA [2021a, p. 435]). Thus, even if FPT results align with the client's financial reporting, more appropriate *external* evidence may still be required to reduce the likelihood of fraud to a reasonable level.

Indeed, external evidence is particularly important in detecting fraudulent financial reporting. Auditing standards stress the importance of considering independent *external* evidence due to its higher reliability and lower susceptibility to management manipulation compared to internal evidence (IAASB [2022a], PCAOB [2024a]). For example, a client's unusually rapid sales growth relative to industry peers may serve as a potential fraud red flag,

warranting further investigation (Brazel, Jones, and Lian [2023], IAASB [2024]). Therefore, vigilance in identifying fraud red flags in external evidence is particularly crucial for maintaining audit quality (IAASB [2022a]). However, we suggest that auditors' use of FPT may trigger an *attribute substitution bias* in their evaluation of the evidence (e.g., Kahneman and Frederick [2002]), potentially impairing their skeptical actions when subsequently confronted with a red flag revealed by external evidence.

2.2 ATTRIBUTE SUBSTITUTION BIAS

Attribute substitution is a heuristic process in which individuals substitute their assessment of a simpler, more readily accessible attribute for their evaluation of a more difficult-to-assess attribute of the same entity (Kahneman and Frederick [2002]). For example, Kadous, Koonce, and Thayer [2012] find that when evaluating the relevance of fair value information—an attribute that is challenging to assess—financial statement users tend to rely on the more easily appraised attribute of measurement reliability. Anderson, Hobson, and Sommerfeldt [2022] observe that investors making investment decisions substitute the assurance on non-GAAP measures—a more accessible attribute—for the actual usefulness of such information, which is less accessible. The results of Choi, Hecht, and Tayler [2012] suggest that decision makers potentially substitute a compensated performance measure (i.e., a more accessible attribute) for strategic constructs (i.e., a less accessible attribute).

While attribute substitution has been studied in financial and managerial accounting contexts, examining the potential presence of attribute substitution bias in the audit setting remains important. Given that evidence sufficiency and evidence appropriateness are two fundamental constructs for audit evidence evaluation, auditing standards clearly define them and distinguish between them. This characteristic of our setting potentially works against

finding the presence of an attribute substitution bias. However, it could have detrimental consequences to audit quality if auditors blur these concepts in their evaluation of evidence.

Auditing standards require auditors to obtain sufficient appropriate audit evidence to support their audit opinion (IAASB [2022a], PCAOB [2024a]). Specifically, auditors are to evaluate their evidence based on two critical aspects: evidence *sufficiency* (quantity) and evidence *appropriateness* (quality). Sufficiency refers to "the measure of the quantity of audit evidence" (IAASB [2022a, p. 24]) and appropriateness is "the measure of the quality of audit evidence" (IAASB [2022a, p. 24]). We posit that auditors' use of FPT may inadvertently lead them to substitute their assessment of evidence *sufficiency* for their judgment of evidence *appropriateness*.

For attribute substitution to influence judgment, three necessary conditions must be met: (1) the existence of a more difficult (i.e., less accessible) attribute that individuals aim to assess; (2) the availability of a more easily evaluated (i.e., more accessible) attribute; and (3) the substitution is unlikely to be challenged by conscious processing (Kahneman and Frederick [2002], Choi, Hecht, and Tayler [2012]). As discussed below, all three conditions are likely satisfied in the context of FPT, suggesting greater potential for attribute substitution. Meanwhile, these conditions are less likely met when traditional sampling methods are employed.

Although auditing standards clearly define evidence sufficiency and appropriateness, their evaluations in practice require significant judgment (e.g., Bamber [1983], Rebele, Heintz, and Briden [1988], Knechel and Messier [1990], Hirst [1994], Altiero, Baudot and Hazgui [2023]). When it comes to evidence appropriateness, auditing standards note that "the quality of audit evidence depends on the relevance and reliability of the information intended to be used as audit evidence as well as the effectiveness of the design of the audit procedures and the auditor's application of those audit procedures" (IAASB [2022a, p. 28]). Therefore, the

assessment of evidence appropriateness is fraught with uncertainties regarding the relevance and reliability of a given piece of evidence, as well as the extent to which the evidence is obtained from effectively designed and performed audit procedures (IAASB [2022a], PCAOB [2021]). Prior research also highlights that auditors consider multiple aspects when evaluating evidence quality, such as the competency of the evidence source (e.g., Bamber [1983], Rebele, Heintz, and Briden [1988], Knechel and Messier [1990], Hirst [1994]).

Evaluating evidence sufficiency can also be challenging. When using a sampling approach, auditors often face uncertainties in determining at what point they have gathered enough evidence (e.g., Altiero, Baudot and Hazgui [2023]). Specifically, Altiero, Baudot and Hazgui [2023] discuss eight potential rules auditors might employ when making these determinations, all of which require substantial professional judgment. For example, auditors typically gauge sample sufficiency by appraising the sample coverage percentage in relation to the full population (e.g., Altiero, Baudot and Hazgui [2023]). In addition, the existence of multiple sampling methods (e.g., random sampling, risk-based sampling, etc.) could further obscure the judgment, making it more difficult to assess sufficiency. In sum, the evaluations of both evidence sufficiency and evidence appropriateness require substantial auditor judgment when using traditional sampling methods. Hence, in sample testing, attribute substitution is relatively unlikely because there is no more easily accessible attribute to substitute, meaning the second necessary condition noted earlier is not met.

In contrast, testing the entire population of client-internal data by employing FPT simplifies the assessment of evidence sufficiency related to a given audit procedure. Unlike sample testing, which examines only a subset of items, FPT scrutinizes 100 percent (i.e., the entire population) of the items comprising a class of transactions or account balance (e.g., IAASB [2022a]). Consequently, FPT eliminates uncertainties surrounding testing coverage and makes the assessment of evidence *sufficiency* straightforward. In other words, auditors

performing FPT can *easily* conclude that sufficient evidence related to a given audit procedure has been obtained, thus satisfying the second necessary condition. Meanwhile, merely expanding the *extent* of the same audit procedures, FPT does not reduce the difficulty of evaluating evidence appropriateness, as the design of the audit procedure and the evidence source remain unchanged. In other words, the evaluation of evidence *appropriateness* remains less accessible in the case of FPT, satisfying the first necessary condition for attribute substitution.

When using FPT, auditors may therefore be inclined to substitute their more accessible assessment of evidence *sufficiency* for their comparatively less accessible evaluation of evidence *appropriateness*. Importantly, this substitution process is unlikely to be consciously challenged in a typical audit setting (the third necessary condition), given various incentives that can cloud auditor conscious processing, such as time pressure and budget constraints (e.g., Kelly, Margheim, and Pattison [1999], Hatfield, Jackson, and Vandervelde [2011], Brazel et al. [2016], Bennett and Hatfield [2017], Lambert et al. [2017], Bhaskar, Hopkins, and Schroeder [2019]). As a result of this substitution, auditors may perceive that the client's financial reporting has been corroborated by seemingly more appropriate evidence. This inflated assessment of evidence appropriateness when employing FPT is likely to reduce auditors' skepticism when they later encounter a fraud red flag related to more appropriate external evidence. Although a reduced need for additional *corroborating* evidence after FPT may be innocuous, being less skeptical in reaction to a *fraud red flag* could potentially hurt audit quality. This leads to our first hypothesis, stated formally:

> *H1*: Auditors using full population testing, compared to sample testing, are less likely to act skeptically when a fraud red flag related to external evidence is subsequently encountered.

2.3 VISUALIZING FULL POPULATION TESTING RESULTS

We further expect that the attribute substitution bias will be more pronounced when the FPT results are presented visually rather than in tabular form. Data analytics techniques, including FPT, frequently employ visualizations like graphs or charts to present data patterns or results (AICPA [2017], Austin et al. [2021], Walker, Brown-Liburd, and Barr-Pulliam [2022], Henry et al. [2023]). Henry et al. [2023] provide evidence that data visualizations are increasingly used in audit tasks, with about 80 percent of their survey respondents indicating their use. Given that auditors often lack the technical expertise to fully engage with emerging technologies during their application (e.g., Walker, Brown-Liburd, and Barr-Pulliam [2022]), they tend to rely on visual outputs from data analytic tools (e.g., Austin et al. [2021]).

Research, in general, supports the benefits of data visualizations (e.g., Benbasat and Dexter [1986], Vessey [1991], Huang, Eades, and Hong [2009], Eberhard [2023]). Using data visualizations enhances auditors' insights into their clients' businesses and data (e.g., KPMG [2015], PwC [2015], Deloitte [2016], Wilsonan and Dennis [2024]). Visualizations also help auditors identify evidence inconsistencies (e.g., Dilla and Raschke [2015], Higginbotham, Nash, and Demeré [2021], Rose et al. [2022], Baaske, Eulerich, and Wood [2023]). For example, Rose et al. [2017] find that when traditional audit procedures are supplemented with big data visualizations, auditors are better able to recognize patterns and incorporate evidence inconsistencies into their judgments. Additionally, Backof, Carpenter, and Thayer [2018] observe that graphical presentations increase auditors' skepticism of management's aggressive assumptions for complex estimates.

However, auditors also face potential challenges when using visualizations (e.g., Chang and Luo [2021]). Given the rapid shift from traditional audit methods to data analytics, auditors may lack experience in effectively assessing and processing data visualizations (e.g., Brown-Liburd, Issa, and Lombardi [2015], Cao, Chychyla, and Stewart [2015], Austin et al.

[2021]). For example, auditors might over-rely on default visualization formats that may not be optimal. Baaske [2021] examines this possibility and finds that when revenue is presented on a monthly basis (default format) rather than weekly (optimal format), auditors make lower risk assessments because a monthly presentation is less likely to reveal improper revenue recognition occurring in the last week of each month. Additionally, the effectiveness of visualizations can vary depending on the audit phase in which they are utilized, whether during risk assessment or substantive testing (Anderson, Hobson, and Peecher [2020]).

Contributing to this line of research, we posit that presenting FPT results in a visualized format could potentially exacerbate FPT's adverse impact on auditors' skepticism. Graphical presentation formats may further facilitate auditors' evidence sufficiency evaluations by easing the processing of visualized FPT results (e.g., Small [1996], Alonso et al. [1998], McBride and Caldara [2013]). For example, auditors can more readily assess their extent of testing by observing the heights of bars in a bar graph, with bars illustrating the testing performed being very tall and equal to the level of the population.

Moreover, visualizing FPT results may reinforce the notion that the attribute substitution is unlikely to be consciously challenged. Research suggests that visualized formats potentially reduce the cognitive resources and attention required to interpret the presented information (e.g., Benbasat and Dexter [1986], Townsend and Kahn [2014], Backof, Carpenter, and Thayer [2018], Eberhard [2023]). Specifically, Rose et al. [2022] discover that bar graphs elicit lower cognitive resource levels and reduce the intensity of cognitive processing. Hence, auditors viewing FPT results visually, compared to those viewing FPT results in a tabular format, are less likely to engage in the conscious processing that could challenge or resist the substitution bias. This leads to our second hypothesis, stated formally:

H2: The negative effect of full population testing on auditors' skeptical actions is exacerbated when the testing results are visualized compared to tabulated.

3. Method

3.1 EXPERIMENTAL DESIGN

To test our hypotheses, we conducted an experiment employing a between-participants design.[4] We manipulated the *Use of FPT* (FPT versus sample testing) and the *Presentation Format* (visualized versus tabulated testing results), resulting in four treatment conditions to which participants were randomly assigned.[5] We adapted the experimental case from Brazel, Leiby, and Schaefer [2022] and Barr-Pulliam et al. [2023] with the authors' consent. Representatives from the Foundation for Auditing Research (FAR) and audit professionals from the participating firms reviewed our instrument to ensure its clarity and realism. Participants completed the experiment online.

3.2 PARTICIPANTS

One hundred and twenty-five audit practitioners in the Netherlands completed our online instrument (mean audit experience = 5.80 years).[6] FAR liaisons within two audit firms assisted us in recruiting participants. At one audit firm, participants completed our online instrument on their laptops at a predetermined time during their in-person training sessions. At the other audit firm, liaisons called for participants via email, allowing them to complete the study during work hours but at a time of their convenience.[7]

---

[4] We obtained Institutional Research Ethical Review Board approval for the experiment used in this study.

[5] Our experiment contained a third randomly assigned independent variable. Specifically, we expected that varying the presentation format of the external information containing the fraud red flag in the *subsequent* year-end analytical procedure task (visualized versus tabulated) would affect the saliency of that inconsistency. We hypothesized that visualizing the external red flag would counteract the negative effect of FPT on auditors' professional skepticism. However, we did not observe the expected interaction pattern. For reasons of parsimony, we collapse this additional independent variable in our analyses. Our primary inference for H1 is robust to including this additional independent variable and its interactions with the other two manipulated independent variables. We do observe a positive main effect of visualizing the external red flag on participants' skeptical actions.

[6] We excluded nine responses due to anomalous completion times. Six participants took excessively long periods (over four hours) while three participants completed the instrument very quickly (less than twelve minutes).

[7] During the data collection process, we were concerned about securing enough participants through the FAR. Consequently, we collected 37 additional responses from students in a part-time Master's in Accounting program at a Dutch university. However, since we eventually received enough experienced participant responses through the FAR and the students were too diverse and inexperienced for the task (e.g., mean audit experience =

Sixty-nine percent of participants reported having experience performing three-way match testing related to sales, which is our context for the FPT and sample testing. Participants' average level of experience with performing FPT was 5.37, measured on an 11-point Likert scale ranging from 0 (None) to 10 (Extensive). Participants' average level of experience with using data visualizations was 5.48 measured on an 11-point Likert scale ranging from 0 (None) to 10 (Extensive). Seventy-eight percent of participants had experience performing a substantive analytical procedure related to sales (our experimental task with the industry growth red flag). There are no significant differences across conditions for participants' demographic characteristics, experience levels, or the data collection methods (i.e., at a predetermined time versus at their convenience), confirming successful randomization. Our primary inferences are also robust to controlling for participants' demographic characteristics, experience levels, and data collection method.

3.3 EXPERIMENTAL PROCEDURES AND INDEPENDENT VARIABLES

Our experiment comprised three stages (see appendix A). In Stage 1, participants received an audit case providing background information about a hypothetical audit client ("Ruiter") and the results of a sales account's interim testing employing client-internal data. They then performed a year-end substantive analytical procedure for that sales account, which involved both internal and external evidence. Stage 2 involved questions aimed at testing the success of our manipulations and capturing process variables. Stage 3 collected measures about participants' audit practice experiences.

After receiving the background information in Stage 1, participants were informed that they had previously performed interim tests of transactions for Ruiter's *Sporting Goods Sales*. Specifically, at interim, they had obtained data from Ruiter's database and performed

1.54 years), we opted not to include them. Although p-values increased, presumably due to noise, our primary inferences are robust to including the Master's students' responses.

two three-way matches for sales transactions from 01/01/2022 to 10/31/2022. All participants learned that the first three-way match compared the *quantity* of matching internal sales orders, sales invoices, and shipping documents. The second three-way match compared the *price* of the respective internal sales order, sales invoice, and Ruiter's master price list.[8]

Participants in the *FPT* condition were informed that their two three-way matches had been performed on the *entire population* of sales transactions from 01/01/2022 to 10/31/2022 (see appendix B.1). In contrast, participants in the *Sample Testing* condition learned that their two three-way matches had been performed on *a sample* of the sales transactions from the same time span (see appendix B.2). Participants in the *Visualized Results* condition were provided with three-way match testing results in bar charts (see appendices B.3 and B.5), while the *Tabulated Results* condition presented participants with three-way match testing results in tables (see appendices B.4 and B.6). We held the final results of the three-way match testing constant across all conditions, such that it always resulted in an immaterial audit adjustment of the same amount (i.e., less than 1% of the entire population).[9]

Note that despite the evidence provided by FPT supporting the client's reported sales balance, the possibility of fraud in the sales account still exists for two main reasons. First, in our case, the FPT was conducted during interim testing, which is common given tight deadlines at year-end (i.e., busy season). Thus, it remains possible that, for example, between 10/31/2022 and year-end, the client could engage in practices like channel stuffing to inflate sales. Second, as typical in audit practice, the FPT in our case relied solely on internal data extracted from the client's system, leaving it susceptible to manipulation by client management. Of course, the same concerns exist for our *Sample Testing* condition as well.

[8] Substantive tests of transactions are typically performed at interim or prior to year-end testing (e.g., Arens et al. [2024]). Our operationalization of the two three-way matches follows the FPT of sales transactions example provided in the AICPA's Guide to Audit Data Analytics, specifically, example 4-2 in appendix C of AICPA [2017].

[9] Participants in the *Sample Testing* condition were informed that they extrapolated the sample testing result to the entire population, including an adjustment for sampling risk. This led to an immaterial adjustment.

Moving from interim to *year-end* testing, consistent with prior studies (e.g., Brazel, Leiby, and Schaefer [2022]), we held budget and time pressure at year-end constant and relatively high across all conditions to obtain a realistic expectation of costs related to skeptical actions. Specifically, all participants were informed that their sales substantive analytical procedures were currently about to go over budget, and that the client's filing deadline was approaching.

Participants then performed a year-end substantive analytical procedure task related to the Ruiter's *Sporting Goods Sales* account. Specifically, they (1) developed an expectation for the sales account, (2) documented how they developed their expectation, and (3) compared their expectation to the recorded balance. Participants were equipped with buttons offering access to a comprehensive set of information to develop their expectation for the sales account. This information included both internal and external evidence spanning the past five years and drawn from sources recommended by auditing standards, including prior-year balances, budgets, industry growth rates, growth in related accounts, and growth in non-financial measures (e.g., AS 2305, PCAOB [2024d], ISA 520, IAASB [2009]).

We embedded an *external evidence inconsistency* in the data provided for the year-end substantive analytical procedure task to all participants, suggesting that the client's current year sales growth exceeded the industry growth rate by approximately 17 percent. This inconsistency mirrors the red flag observed for fraudulent firms by Brazel, Jones, and Lian [2023]. To clarify, this red flag was incorporated in the separate information used for the *year-end* substantive analytical procedure, not the data previously described in the *interim* three-way matches. All other data for the year-end substantive analytical procedure were consistent with the client's reported sales balance. Based on the provided data and information, participants proceeded to develop an expectation for the sales account and determined whether additional testing would be necessary as part of the analytical procedure.

3.4 DEPENDENT VARIABLE

Participants were presented with a choice to either perform additional work related to the analytical procedure or decide that no additional work related to the analytical procedure was necessary. For those who chose to perform additional work, we inquired about the specific tasks they would undertake and/or the questions they would pose to client management. Regardless of whether they deemed additional work necessary, all participants were asked if they had anything to communicate to their audit manager. These actions collectively represent the initial skeptical actions that auditors would perform before proceeding to subsequent, more costly skeptical actions, such as expanding substantive testing and collecting additional evidence. Moreover, Brazel et al. [2016] demonstrate that there is variation in the extent to which auditors first inform their manager or choose to directly question client management. Therefore, we assigned a *Skeptical Actions* score of "1" if participants expressed their intent to test or inquire about the industry growth red flag with client management and/or communicate the industry growth red flag to their audit manager, and "0" otherwise.[10]

4. Results

4.1 MANIPULATION CHECKS

To confirm our successful manipulation of *FPT*, participants were asked to indicate the perceived sufficiency of the sales transactions tested in the two three-way matches at interim,

[10] Participants performed an additional task after the substantive analytical procedure. Specifically, they were tasked with determining the sample size and related budget for tracing a sample of Ruiter's open accounts receivable at the year-end to post-year-end cash receipts listed on Ruiter's bank statements. However, with hindsight, we do not believe this task was meaningful, because participants completed it without knowing the outcome of their first task (i.e., the analytical procedure). In particular, those who had chosen to address the fraud red flag had no way of knowing whether management could or could not reasonably explain the inconsistency. Further, participants would then need to determine if additional evidence corroborated management's explanation. Even though we believe it is difficult to adequately interpret participants' recorded sample sizes and budgets for the second task, we briefly report the respective results: Participants chose to sample 7.95 percent of accounts receivable in the FPT condition versus 7.10 percent in the Sample Testing condition and budgeted 9.15 hours on that testing in the FPT condition versus 8.47 hours in the Sample Testing condition. While significant, these differences are unlikely to generate a significant practical impact (e.g., 40 additional minutes of testing).

utilizing an 11-point Likert scale ranging from 0 (Not at all sufficient) to 10 (Completely sufficient). Participants in the *FPT* condition rated the evidence sufficiency significantly higher than those in the *Sample Testing* condition (7.69 vs. 6.75, t_{123} = 2.51, p = 0.01).[11] These results affirm that participants understood and internalized our *FPT* versus *Sample Testing* manipulation.

We next present evidence that participants in the *Visualized Results* condition accurately perceived the three-way match testing results as visualized, while those in the *Tabulated Results* condition correctly identified the results as tabulated. Out of 53 participants in the *Visualized Results* condition, 47 (i.e. 89 percent) correctly indicated that the three-way matches were displayed in visualizations. Similarly, out of the 72 participants in the *Tabulated Results* condition, 63 (i.e. 88 percent) correctly identified that the two three-way matches were presented in tables. We conclude that our manipulation of the presentation format was also successful.[12]

4.2 TEST OF H1: FULL POPULATION TESTING VERSUS SAMPLE TESTING

The observed pattern of our results is presented in figure 1. Table 1, panel A reports descriptive statistics for participants' *Skeptical Actions* related to the external fraud red flag. Table 1, panel B presents the logistic regression results with participants' *Skeptical Actions* as the dependent variable. Auditors are less likely to inquire of client management about the fraud red flag and/or communicate the red flag to their audit manager when using FPT compared to sample testing (0.48 vs. 0.65, β = -0.73, Wald $\chi^2_{(1)}$ = 3.85, p = 0.03, one-tailed). The logistic regression coefficient indicates that auditors in *FPT* condition are 52 percent less

[11] We report two-tailed p-values unless otherwise noted.
[12] Our primary inferences are robust to excluding responses that do not pass the manipulation check for presentation format.

likely to address the industry growth red flag with client management or their auditor manager compared to auditors in the *Sample Testing* condition.[13] Therefore, H1 is supported.

4.3 TEST OF H2: VISUALIZING VERSUS TABULATING TESTING RESULTS

H2 predicts that the negative effect of *FPT* on auditors' *Skeptical Actions* is exacerbated when the testing results are visualized versus tabulated (*Presentation Format*). As shown in figure 1 and table 1, panel B, although the direction of the results is consistent with our prediction, the interaction of *Use of FPT* and *Presentation Format* is not significant. H2 is therefore not supported.

4.4 SUPPLEMENTAL ANALYSES

4.4.1. Substituting Sufficiency for Appropriateness. With H1, our theory posits that when auditors employ FPT as opposed to sample testing, they will substitute their judgment of evidence sufficiency for their evaluation of evidence appropriateness. In other words, when using FPT compared to sample testing, auditors are expected to perceive a higher level of internal evidence sufficiency and consequently an inflated sense of internal evidence appropriateness. Post-experiment, we captured participants' evaluations of evidence sufficiency (*Sufficiency*) by asking about them to rate the sufficiency of the sales transactions tested in the two three-way matches at interim on an 11-point Likert scale, ranging from 0 (Not at all sufficient) to 10 (Completely sufficient). We also measured participants' evaluations of evidence appropriateness (*Appropriateness*) by asking them to rate the appropriateness of the evidence obtained from the two three-way match tests on an 11-point Likert scale, ranging from 0 (Not at all appropriate) to 10 (Highly appropriate).

We first examine the evidence sufficiency and appropriateness substitution process using Hayes' [2018] PROCESS model 4 in SPSS. Using 5,000 bootstrap resamples with

[13] To transform the logistic regression coefficient for the main effect of FPT (i.e., -0.73), which represents the logarithm of the odds ratio, into percentage change, we exponentiated it (i.e., $e^{-0.73} = 0.48$) and then subtracted the calculated odds ratio from 1 (i.e., 1 – 0.48 = 0.52).

replacement, we estimate 90% confidence intervals for the indirect effect, with significant mediations indicated by intervals excluding zero. Consistent with our theory, we find that participants evaluate the internally generated evidence through FPT compared to sample testing as more *sufficient* (a_1 = 0.94, p = 0.01), subsequently inflating their evaluation of the internal evidence's *appropriateness* (b_1 = 0.30, p < 0.01) as presented in figure 2. The significant indirect effect ($a_1 * b_1$ = 0.28, SE = 0.17; 90% LLCI = 0.06, 90% ULCI = 0.61) provides evidence of auditors substituting their evaluation of evidence sufficiency for evidence appropriateness.

More importantly, FPT itself does not directly increase auditors' evidence appropriateness evaluation. As presented in figure 2, there is no direct effect of FPT on *Appropriateness* (a_2 = 0.30, p = 0.41). Consistent with this observation, Hayes' [2018] PROCESS model 6 in figure 3 shows that the path from *Use of FPT* to *Skeptical Actions* through *Appropriateness* alone is not significant ($a_2 * b_3$ = 0.01, SE = 0.05; 90% LLCI = -0.05, 90% ULCI = 0.10). These results suggest that appropriateness perceptions are inflated because auditors perceive that FPT increases sufficiency, not because they believe FPT increases appropriateness independent of their sufficiency assessments.

Furthermore, the increased *Sufficiency* alone does not mediate the path from *Use of FPT* to *Skeptical Actions* in figure 3 ($a_1 * b_2$ = 0.07, SE = 0.11; 90% LLCI = -0.09, 90% ULCI = 0.25), indicating that the negative effect of FPT on auditors' skeptical actions is not due to their increased sufficiency assessment independent of the inflated appropriateness. These observations corroborate our theoretical mechanism of auditors' substituting the attribute of sufficiency for appropriateness when employing FPT.

Figure 3 further reveals that the inflated *Appropriateness* evaluation does not affect *Skeptical Actions* (b_3 = 0.03, p = 0.78). In the exploratory analysis that follows, we examine

whether varying levels of experience with FPT might explain the absence of a mediation effect on skeptical actions.

4.4.2. The Role of Auditors' Experience with FPT. We further investigate whether participants' *personal experience* with using FPT affects our findings. Participants' overall level of experience with performing full population testing (*FPT Experience*) was 5.37, utilizing an 11-point Likert scale ranging from 0 (None) to 10 (Extensive). We observe that there is substantial variance in participants' experience with using FPT (Standard Deviation = 2.80, Coefficient of Variation = 52.14%). Auditors with and without a track record of using FPT in practice likely hold differing beliefs about how FPT impacts the evidence set used to support the audit opinion.

Using Hayes' [2018] PROCESS model 89, we find that the indirect effect of *FPT* on *Skeptical Actions* through the serial mediation of *Sufficiency* and *Appropriateness* is indeed moderated by *FPT Experience*. Results in figure 4 reveal a negative coefficient for the interaction between *Appropriateness* and *FPT Experience* (b_5 = -0.07, p = 0.08) on *Skeptical Actions*. This result indicates that more experience with using FPT negatively influences their skeptical actions resulting from an inflated sense of evidence appropriateness after using FPT (vs. sampling). The results also present a significant moderated mediation index (index = -0.02, SE = 0.02, 90% LLCI = -0.06, 90% ULCI = -0.00).[14] These findings confirm that the negative indirect effect of *FPT* on *Skeptical Actions* through *Sufficiency* and *Appropriateness* becomes *stronger* as auditors' FPT experience increases. Given the continued rise of FPT in practice and that auditors' experiences with FPT will undoubtedly increase in the future, this finding of the substitution effect increasing with FPT experience is concerning. It is also consistent with prior research concluding that experience or knowledge might not always enhance skepticism (e.g., Nelson [2009]). For instance, more experienced auditors are more

[14] The 90% ULCI before rounding is -0.0005.

likely to assume non-error explanations for audit findings (e.g., Kaplan, Moeckel, and Williams [1992], Solomon, Shields, and Whittington [1999]).

4.4.3. Ruling Out Confidence that a Material Misstatement Does Not Exist. We examine whether the observed effect of *FPT* can be explained by differences in auditors' confidence that there was no material misstatement in the client's sales transactions at interim. Since FPT examines the full population and can identify all potential errors, auditors using FPT, compared to sample testing, may have held a higher level of confidence in the absence of a material misstatement in the client's reporting at interim. However, such confidence would not justify the failure to act skeptically in the presence of a subsequent external fraud red flag.

We post-experimentally asked participants to rate their confidence in concluding that the evidence obtained from the two three-way matches supported that there was no material misstatement in the client's sales transactions. Responses were obtained on an 11-point Likert scale, ranging from 0 (Not at all Confident) to 10 (Highly Confident). Results (untabulated) show no differences between the *FPT* and *Sample Testing* conditions for this measure. Further, using the Hayes [2018] PROCESS model 4 in SPSS, we find no evidence that confidence mediates the effect of *FPT* versus *Sample Testing* on participants' skeptical actions. Therefore, we rule out the possibility that higher confidence derived from the FPT versus sample testing drove our results.

Equal levels of confidence in the absence of a material misstatement across *FPT* and *Sample Testing* conditions are potentially explained by the fact that, in our experiment, sampling risk was incorporated into the sampling extrapolation and all participants received the same final results from the interim tests (i.e., an immaterial adjustment of the sales account). They may also reflect audit firms' beliefs that while data analytics can improve audit efficiency and offer deeper insights, it does not necessarily alter the level of assurance (i.e., reasonable assurance) provided by an audit (e.g., IAASB [2016]). FPT versus sample

testing inflating auditors' perceptions of evidence appropriateness, but not their assessed level of assurance (i.e., confidence that a material misstatement does not exist), potentially implies that the attribute substitution process is subconscious.

5. Conclusion

The emergence of data analytics has provided auditors with the capability to test entire populations of client-internal data, moving beyond traditional sample-based testing. Although FPT offers numerous benefits, we demonstrate that it inadvertently impairs auditors' skeptical actions in response to the subsequent revelation of a fraud red flag. Specifically, relying on FPT results triggers an attribute substitution bias, where auditors deem the evidence generated from the client's information system as not only more sufficient but also more appropriate than that obtained through sample testing. Consequently, auditors using FPT become less skeptical in the face of a fraud red flag subsequently revealed by more appropriate (external) evidence. This negative effect of FPT on auditors' skeptical actions is not dependent on how the FPT results are presented—through visualizations or tables. Overall, this study contributes to the growing literature on the effects of audit technologies on auditors' professional skepticism by revealing an unintended consequence of FPT on auditors' evidence evaluation and their subsequent skeptical actions (e.g., Rose et al. [2017], Anderson, Hobson, and Peecher [2020], Baaske, Eulerich, and Wood [2023]).

The importance of external evidence has been increasingly recognized in auditing, both in the academic literature and among regulators and standard setters (e.g., PCAOB [2021], Brazel, Jones, and Lian [2023]). Technological advancements, such as the use of artificial intelligence, are expected to enhance access to a broader range of information from external sources (e.g., PCAOB [2021], IAASB [2022b]). However, the results of this study suggest that the increased use of FPT on internal evidence may inadvertently lead auditors to underutilize such external evidence, despite its potential to improve audit quality. These

insights are valuable for standard-setters and regulators attempting to understand how auditors integrate and leverage external evidence in light of advancing audit technologies (e.g., PCAOB [2021, 2024c]).

More broadly, the current study significantly contributes to the literature on auditors' evaluations of audit evidence. It builds upon prior research that finds that the failure to obtain sufficient appropriate audit evidence is among the leading causes of audit deficiencies (e.g., Beasley, Carcello, and Hermanson [2001], Church and Shefchik [2012], Beasley et al. [2013], Rajgopal, Srinivasan, and Zheng [2021]). Auditing standards clearly define and distinguish between evidence sufficiency and appropriateness, emphasizing that evidence *quantity* cannot compensate for low evidence *quality* (IAASB [2022a], PCAOB [2024a]). However, it remains unclear how auditors evaluate evidence in practice given auditors' judgments are potentially susceptible to biases (e.g., Hackenbrack [1992], McMillan and White [1993], Joe, Vandervelde, and Wu [2017], Lambert and Peytcheva [2020]). Although auditing standards establish a *unidirectional* relationship between evidence appropriateness and evidence sufficiency, this study introduces the concept of attribute substitution—a theory not previously explored in the auditing context—to explain why auditors may mistakenly perceive their relationship as *bidirectional* and regard more sufficient yet lower-quality evidence as more appropriate.

The phenomenon of attribute substitution bias, as highlighted in this study, may also shed light on auditors' tendency to ineffectively respond to heightened fraud risk. Rather than making effective modifications to their audit procedures or *nature* of their testing, auditors often resort to merely expanding the *extent* of the same audit procedures, such as increasing planned sample sizes and budgeted hours (e.g., Hammersley [2011], Hammersley, Johnstone, and Kadous [2011]). That is, auditors appear to equate the increased extent of the same audit procedure with the enhanced evidence appropriateness needed to mitigate higher levels of

fraud risk. To address this issue, audit firms should consider modifying their quality control systems and training programs to ensure auditors better calibrate their reliance on both internal and external evidence, given the level of detection risk. Further, training programs could remind auditors that, despite advances in data analytics that will greatly increase the sufficiency of evidence examined, auditing standards clearly define and distinguish evidence sufficiency and appropriateness, and that audit quality can be compromised if auditors blur those two concepts when using advanced technologies.

While our experiment operationalized FPT by means of increasing the volume of client-*internal* data as is typical in practice, the detrimental effect of FPT triggering attribute substitution bias on auditors' skepticism extends beyond this specific context. Our finding generalizes to FPT incorporating any evidence lacking appropriateness. Moreover, we believe that the underlying mechanism of attribute substitution in auditors' evidence evaluation processes and its negative effect on auditors' skepticism extends beyond the specific technique (i.e., FPT) and source of evidence (i.e., external evidence) examined in this study. Accessible characteristics of other technologies could also trigger the substitution process. For example, auditors could substitute an artificial intelligence tool's humanization characteristic for its reliability, evaluating the evidence from that tool as more reliable than otherwise would be determined. Auditors should be aware of or alerted to the potential existence of attribute substitution bias when using not only FPT but also other technologies. Future research could examine other audit settings where the bias exists, as well as potential remedies for the bias.

Although we do not find support for the expected effect of visualizations in this study, future research could further explore the effects of visualizations on auditors' professional skepticism. For example, do the effects of visualizations depend on their revelation of red flags or not? Visualizations revealing red flags may benefit auditors' professional skepticism,

while those revealing no red flags may unintendedly distract auditors' attention from subsequent red flags revealed in other substantive testing, eventually leading to a negative effect on auditors' professional skepticism.

REFERENCES

ALONSO, D. L.; A. ROSE, C. PLAISANT, and K. L. NORMAN. "Viewing Personal History Records: A Comparison of Tabular Format and Graphical Presentation Using LifeLines. *Behaviour & Information Technology* 17 (1998): 249-62.

ALTIERO, E. C.; L. BAUDOT; and M. HAZGUI. "Are We Good? Or Do We Need to Keep Going? Evidence Sufficiency in Assurance Engagements." 2023. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4411699.

AMATO, N. "Audit Evolution in Action: Part 2." *Journal of Accountancy*, June 29 (2021). Available at https://www.journalofaccountancy.com/podcast/audit-evolution.html.

AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS (AICPA). *Guide to Data Analytics*. Durham, NC: AICPA, 2017.

AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS (AICPA). *AU-C Section 500 Audit Evidence*. Durham, NC: AICPA, 2021a.

AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS (AICPA). *AU-C Section 240 Consideration of Fraud in a Financial Statement Audit*. Durham, NC: AICPA, 2021b.

ANDERSON, S. B.; J. L. HOBSON; and M. E. PEECHER. "The Joint Effects of Rich Data Visualization and Audit Procedure Categorization on Auditor Judgment." 2020. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3737234.

ANDERSON, S. B.; J. L. HOBSON; and R. D. SOMMERFELDT. "Auditing Non-GAAP Measures: Signaling More Than Intended." *Contemporary Accounting Research* 39 (2022): 577-606.

ARENS, A. A.; R. J. ELDER; M. S. BEASLEY; and C. E. HOGAN. *Auditing & Assurance Services: An Integrated Approach, 18th Edition*. Hoboken, NJ: Pearson Education, Inc., 2024.

AUSTIN, A. A.; T. D. CARPENTER; M. H. CHRIST; and C. S. NIELSON. "The Data Analytics Journey: Interactions among Auditors, Managers, Regulation, and Technology." *Contemporary Accounting Research* 38 (2021): 1888-1924.

BAASKE, B. N.; M. EULERICH; and D. A. WOOD. "Improving Audit Quality with Data Analytics: The Importance of Spatial Abilities and Feedback in Anomaly Identification." 2023. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4477590.

BAASKE, R. N. "The Use of Data Analytic Visualizations to Inform the Audit Risk Assessment: The Impact of Initial Visualization Form and Documentation Focus." 2021. Available at https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=10264&context=etd.

BACKOF, A. G.; T. D. CARPENTER; and J. THAYER. "Auditing Complex Estimates: How Do Construal Level and Evidence Formatting Impact Auditors' Consideration of Inconsistent Evidence?" *Contemporary Accounting Research* 35 (2018): 1798-815.

BAMBER, E. M. "Expert Judgment in the Audit Team: A Source Reliability Approach." *Journal of Accounting Research* 21 (1983): 396-412.

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. L. BROWN-LIBURD; and K. A. SANDERSON. "The Effects of the Internal Control Opinion and Use of Audit Data Analytics on Perceptions of Audit Quality, Assurance, and Auditor Negligence." *Auditing: A Journal of Practice and Theory* 41 (2021): 25-48.

BARR-PULLIAM, D.; J. F. BRAZEL; J. MCCALLEN; and K. WALKER. "Data Analytics and Skeptical Actions: The Countervailing Effects of False Positives and Consistent

Rewards for Skepticism." 2023. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3537180.
BDO. "The Future of the Audit in 5 Predictions." February 2022. Available at https://www.bdo.com/insights/assurance/the-future-of-the-audit-in-5-predictions.
BEASLEY, M. S.; J. V. CARCELLO; and D. R. HERMANSON. "Top 10 Audit Deficiencies." *Journal of Accountancy* 19 (2001): 63.
BEASLEY, M. S.; J. V. CARCELLO; D. R. HERMANSON; and T. NEAL. *An Analysis of Alleged Auditor Deficiencies in SEC Fraud Investigations: 1998–2010.* Washington, DC: Center for Audit Quality (CAQ), 2013.
BENBASAT, I., and A. S. DEXTER. "An Investigation of the Effectiveness of Color and Graphical Information Presentation under Varying Time Constraints." *MIS Quarterly* 10 (1986): 59-83.
BENNETT, G. B., and R. C. HATFIELD. "Do Approaching Deadlines Influence Auditors' Materiality Assessments?" *Auditing: A Journal of Practice and Theory* 36 (2017): 29-48.
BHASKAR, L. S.; P. E. HOPKINS; and J. H. SCHROEDER. "An Investigation of Auditors' Judgments When Companies Release Earnings Before Audit Completion." *Journal of Accounting Research* 57 (2019): 355-390.
BRAZEL, J. F.; J. LEIBY; and T. R. SCHAEFER. "Do Rewards Encourage Professional Skepticism? It Depends." *The Accounting Review* 97 (2022): 131-54.
BRAZEL, J. F.; K. L. JONES; and Q. LIAN. "Auditor Use of Benchmarks to Assess Fraud Risk: The Case for Industry Data." 2023. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3591263.
BRAZEL, J. F.; S. B. JACKSON; T. J. SCHAEFER; and B. W. STEWART. "The Outcome Effect and Professional Skepticism." *The Accounting Review* 91 (2016): 1577-99.
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." *The Accounting Review* 97 (2015): 131-154.
CAO, M.; R. CHYCHYLA; and T. STEWART. "Big Data Analytics in Financial Statement Audits." *Auditing: A Journal of Practice & Theory* 29 (2015): 423-29.
CHANG, C. J., and Y. LUO. "Data Visualization and Cognitive Biases in Audits." *Managerial Auditing Journal* 36 (2021): 1-16.
CHOI, J.; G. W. HECHT; and W. B. TAYLER. "Lost in Translation: The Effects of Incentive Compensation on Strategy Surrogation." *The Accounting Review* 87 (2012): 1135-63.
CHURCH, B. K., and L. B. SHEFCHIK. "PCAOB Inspections and Large Accounting Firms." *Accounting Horizons* 26 (2012): 43-63.
DELOITTE. "The Power of Advanced Audit Analytics: Everywhere Analytics." 2016. Available at https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-da-advanced-audit-analytics.pdf.
DILLA, W. N., and R. L. RASCHKE. "Data Visualization for Fraud Detection: Practice Implications and A Call for Future Research." *International Journal of Accounting Information Systems* 16 (2015): 1-22.
EBERHARD, K. "The Effects of Visualization on Judgment and Decision-Making: A Systematic Literature Review." *Management Review Quarterly* 73 (2023): 167-214.
ERNST & YOUNG (EY). "How Can Data and Technology Help Deliver a High-Quality Audit?" February 2023. Available at https://www.ey.com/en_us/digital-audit/data-and-tech-deliver-high-quality-audit.
FINANCIAL REPORTING COUNCIL (FRC). "Technological Resources: Using Technology to Enhance Audit Quality" March 2020. Available at https://media.frc.org.uk/documents/Data_Analytics_Thematic_Mar_20.pdf.

FREIMAN, J. W.; Y. KIM; and M. A. VASARHELYI. "Full Population Testing: Applying Multidimensional Audit Data Sampling (MADS) to General Ledger Data Auditing." *International Journal of Accounting Information Systems* 46 (2022): 100573.
GLOVER, S. M.; M. H. TAYLOR; and Y. J. WU. "Mind the Gap: Why Do Experts Have Differences of Opinion regarding the Sufficiency of Audit Evidence Supporting Complex Fair Value Measurements?" *Contemporary Accounting Research* 36 (2019): 1417-60.
HACKENBRACK, K. "Implications of Seemingly Irrelevant Evidence in Audit Judgment." *Journal of Accounting Research* 30 (1992): 126-36.
HAMMERSLEY, J. S. "A Review and Model of Auditor Judgments in Fraud-Related Planning Tasks." *Auditing: A Journal of Practice & Theory* 30 (2011): 101-28.
HAMMERSLEY, J. S.; K. M. JOHNSTONE; and K. KADOUS. "How Do Audit Seniors Respond to Heightened Fraud Risk?" *Auditing: A Journal of Practice & Theory* 30 (2011): 81-101.
HATFIELD, R. C.; S. B. JACKSON; and S. D. VANDERVELDE. "The Effects of Prior Auditor Involvement and Client Pressure on Proposed Audit Adjustments." *Behavioral Research in Accounting* 23 (2011): 117-30.
HAYES, A. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, Second Edition*. New York, NY: Guilford Press, 2018.
HENRY, T. F.; M. P. HOLTZMAN; D. A. ROSENTHAL; and R. R. WEITZ. "The Use of Data Analytics in Auditing: Searching for Reality within the Hype." *Journal of Accountancy*, September (2023). Available at https://www.cpajournal.com/2023/09/13/the-use-of-data-analytics-in-auditing.
HIGGINBOTHAM, N.; L. NASH; and W. DEMERÉ. "Making Audits More Effective through Data Visualization." *Journal of Accountancy*, May 1 (2021). Available at https://www.journalofaccountancy.com/issues/2021/may/make-audits-more-effective-through-data-visualization.html.
HIRST, D. E. "Auditors' Sensitivity to Source Reliability." *Journal of Accounting Research* 32 (1994): 113-26.
HOBSON, J. L.; W. J. MAYEW; M. E. PEECHER; and M. VENKATACHALAM. "Improving Experienced Auditors' Detection of Deception in CEO Narratives." *Journal of Accounting Research* 55 (2017): 1137-66.
HOOGDUIN, L. A.; T. W. HALL; J. J. TSAY; and B. J. PIERCE. "Does Systematic Selection Lead to Unreliable Risk Assessments in Monetary-Unit Sampling Applications?" *Auditing: A Journal of Practice & Theory* 34 (2015): 85-107.
HUANG, F.; W. G. NO; M. A. VASARHELYI; and Z. YAN. "Audit Data Analytics, Machine Learning, and Full Population Testing." *The Journal of Finance and Data Science* 8 (2022): 138-44.
HUANG, W.; P. EADES; and S. HONG. "Measuring the Effectiveness of Graph Visualizations: A Cognitive Load Perspective." *Information Visualization* 8 (2009): 139-152.
INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Analytical Procedures. ISA 520*. New York, NY: International Federation of Accountants (IFAC), 2009.
INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). "Exploring the Growing Use of Technology in the Audit, with a Focus on Data Analytics." 2016. Available at https://www.iaasb.org/publications/exploring-growing-use-technology-audit-focus-data-analytics.
INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Identifying and Assessing the Risks of Material Misstatement. ISA 315 Revised.* New York, NY: International Federation of Accountants (IFAC), 2019.

INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *Audit Evidence. Proposed ISA 500 Revised*. New York, NY: International Federation of Accountants (IFAC), 2022a.

INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). "Disruptive Technologies Roundtable Summary." 2022b. Available at https://www.iaasb.org/publications/disruptive-technologies-roundtable-summary.

INTERNATIONAL AUDITING AND ASSURANCE STANDARDS BOARD (IAASB). *The Auditor's Responsibilities Relating to Fraud in An Audit of Financial Statements. Proposed ISA 240 Revised.* New York, NY: International Federation of Accountants (IFAC), 2024.

JOE, J. R.; S. D. VANDERVELDE; and Y. J. WU. "Use of High Quantification Evidence in Fair Value Audits: Do Auditors Stay in Their Comfort Zone?" *The Accounting Review* 92 (2017): 89-116.

JOHNSON, R. N. and L. D. WILEY. *Auditing A Practical Approach with Data Analytics*. Hoboken, NJ: John Wiley and Sons, Inc., 2022.

KADOUS, K.; L. KOONCE; and J. M. THAYER. "Do financial statement users judge relevance based on properties of reliability?" *The Accounting Review* 87 (2012): 1335-56.

KAHNEMAN, D., and S. FREDERICK. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman. Cambridge, UK: Cambridge University Press, 2002, 49-81.

KAPLAN, S. E.; C. MOECKEL; and J. D. WILLIAMS. "Auditors' hypothesis plausibility assessments in analytical review setting." *Auditing: A Journal of Practice & Theory* 11 (1992): 50-65.

KELLY, T.; L. MARGHEIM; and D. PATTISON. "Survey on the differential effects of time deadline pressure versus time budget pressure on auditor behavior." *Journal of Applied Business Research* 15 (1999): 117-128.

KNECHEL, W. R., and W. F. MESSIER. "Sequential Auditor Decision Making: Information Search and Evidence Evaluation." *Contemporary Accounting Research* 6 (1990): 386-406.

KPMG. "Audit Data and Analytics: Unlocking the Value of Audit." February 2015. Available at https://kpmg.com/xx/en/home/insights/2015/02/audit-data-analytics-unlocking-value-of-audit.html.

LAMBERT, T. A., and M. PEYTCHEVA. "When Is the Averaging Effect Present in Auditor Judgments?" *Contemporary Accounting Research* 37 (2020): 277-296.

LAMBERT, T. A.; K. L. JONES; J. F. BRAZEL; and D. S. SHOWALTER. "Audit time pressure and earnings quality: An examination of accelerated filings." *Accounting, Organizations and Society* 58 (2017): 50-66.

MCALLISTER, M.; A. D. BLAY; and K. KADOUS. "Fraud Brainstorming Group Composition: The Persuasive Power of a Skeptical Minority. *The Accounting Review* 96 (2021): 431-48.

MCBRIDE, M., and M. CALDARA. "The Efficacy of Tables versus Graphs in Disrupting Dark Networks: An Experimental Study." *Social Networks* 35 (2013): 406-22.

MCMILLAN, J. J., and R. A. WHITE. "Auditors' Belief Revisions and Evidence Search: The Effect of Hypothesis Frame, Confirmation Bias, and Professional Skepticism." *The Accounting Review* 68 (1993): 443-65.

NELSON, M. W. A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice & Theory* 28 (2009): 1-34.

PREACHER, K., and A. HAYES. "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models." *Behavior Research Methods* 40 (2008): 879-91.

PRICEWATERHOUSECOOPERS (PwC). "How Auditors Combine Tech Know-How and Finance Skills to Drive Innovation." October 2022. Available at https://www.pwc.com/us/en/tech-effect/automation/audit-technology-and-digital-skills.html.

PRICEWATERHOUSECOOPERS (PwC). "PwC Transforms Audit Practice with Data Analytics." November 2015. Available at https://www.accountingtoday.com/opinion/pwc-transforms-audit-practice-with-data-analytics.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). "Staff Guidance – Insights for Auditors Evaluating Relevance and Reliability of Audit Evidence Obtained from External Sources." 2021. Available at https://pcaob-assets.azureedge.net/pcaob-dev/docs/default-source/standards/documents/evaluating-relevance-and-reliability-of-audit-evidence-obtained-from-external-sources.pdf?sfvrsn=48b638b6.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *The Auditor's Use of Confirmation, and Other Amendments to PCAOB Standards. PCAOB Release No. 2023-008.* 2023. Available at https://assets.pcaobus.org/pcaob-dev/docs/default-source/rulemaking/docket_028/2023-008_confirmation-adopting-release.pdf?sfvrsn=e18cef74_4.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Amendments Related to Aspects of Designing and Performing Audit Procedures that Involve Technology-Assisted Analysis of Information in Electronic Form. PCAOB Release No. 2024-007*. 2024b. Available at https://assets.pcaobus.org/pcaob-dev/docs/default-source/rulemaking/docket-052/2024-007-adoptingrelease.pdf.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Auditing Standard 1105: Audit Evidence (effective for fiscal years beginning on or after 12/15/2025)*. 2024a. Available at https://pcaobus.org/oversight/standards/auditing-standards/details/as-1105--audit-evidence-(effective-for-fiscal-years-beginning-on-or-after-12-15-2025).

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Auditing Standard* 2305: *Substantive Analytical Procedures (effective for fiscal years beginning on or after 12/15/2024)*. 2024d. Available at https://pcaobus.org/oversight/standards/auditing-standards/details/as-2305--substantive-analytical-procedures-(effective-for-fiscal-years-beginning-on-or-after-12-15-2024).

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *SEC Filing Form 19b-4. File No. PCAOB-2024-03*. 2024c. Available at https://www.sec.gov/files/rules/pcaob/2024/34-100430.pdf.

RAJGOPAL, S.; S. SRINIVASAN; and X. ZHENG. "Measuring audit quality." *Review of Accounting Studies* 26 (2021): 559-619.

REBELE, J. E.; J. A. HEINTZ; and G. E. BRIDEN. "Independent Auditor Sensitivity to Evidence Reliability." *Auditing: A Journal of Practice & Theory* 8 (1988): 43-52.

ROSE, A. M.; J. M. ROSE; K. A. SANDERSON; and J. C. THIBODEAU. "When should audit firms introduce analyses of big data into the audit process?" *Journal of Information Systems* 31 (2017): 81-99.

ROSE, A. M.; J. M. ROSE; K. ROTARU; K. A. SANDERSON; and J. C. THIBODEAU. "Effects of data visualization choices on psychophysiological responses, judgment, and audit quality." *Journal of Information Systems* 36 (2022): 53-79.

SMALL, D. Navigating Large Bodies of Text. *IBM Systems Journal* 35 (1996): 514-25.

SOLOMON, I.; M. D. SHIELDS; and O. R. WHITTINGTON. What do industry-specialist auditors know? *Journal of Accounting Research* 37 (1999): 191–208.

TOWNSEND, C., and B. E. KAHN. "The "Visual Preference Heuristic": The Influence of Visual versus Verbal Depiction on Assortment Processing, Perceived Variety, and Choice Overload." *Journal of Consumer Research* 40 (2014): 993-1015.

TYSIAC, K. “Embracing Technology in the Audit.” *Journal of Accountancy*, February 1 (2022). Available at https://www.journalofaccountancy.com/issues/2022/feb/embracing-technology-audit.htm.

VESSEY, I. “Cognitive Fit: A Theory Based Analysis of the Graphs versus Tables Literature.” *Decision Sciences* 22 (1991): 219-40.

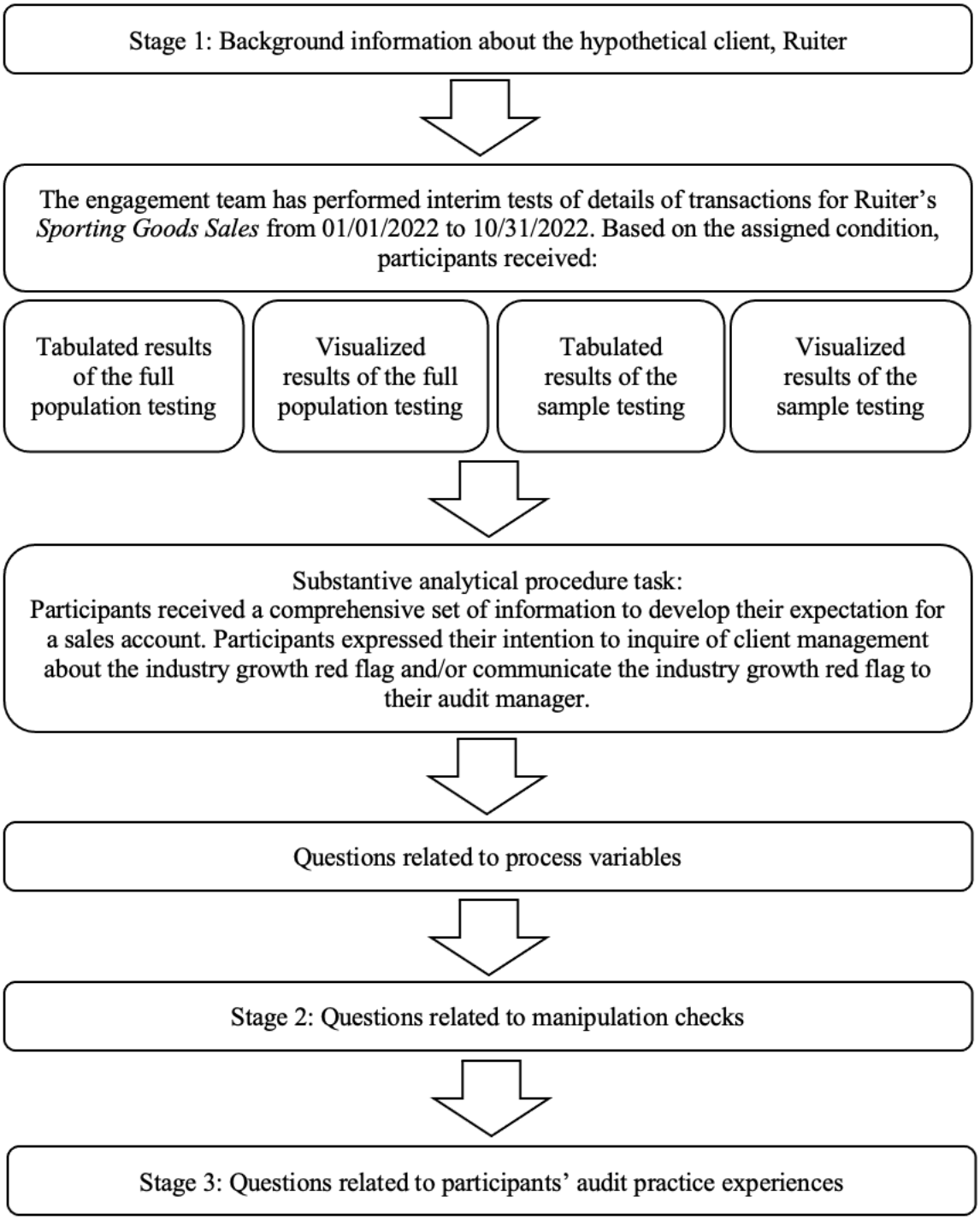
WALKER, K.; H. BROWN-LIBURD; and D. BARR-PULLIAM. “Embracing a Paradoxical Environment to Promote Technological Advancements in Auditing: Perspectives from Auditors in the Field.” 2022. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4286205.

WILSONAN, B., and A. DENNIS. “Data Analytics and Visualization in the Audit.” *Journal of Accountancy*, March 1 (2024). Available at https://www.journalofaccountancy.com/issues/2024/mar/data-analytics-and-visualization-in-the-audit.html.

WURST, J.; J. NETER; and J. GODFREY. “Effectiveness of Rectification in Audit Sampling.” *The Accounting Review* 66 (1991), 333-46.

APPENDIX A

Summary of Experimental Procedures



APPENDIX B

Experiment Excerpts

B.1 FULL POPULATION TESTING CONDITION

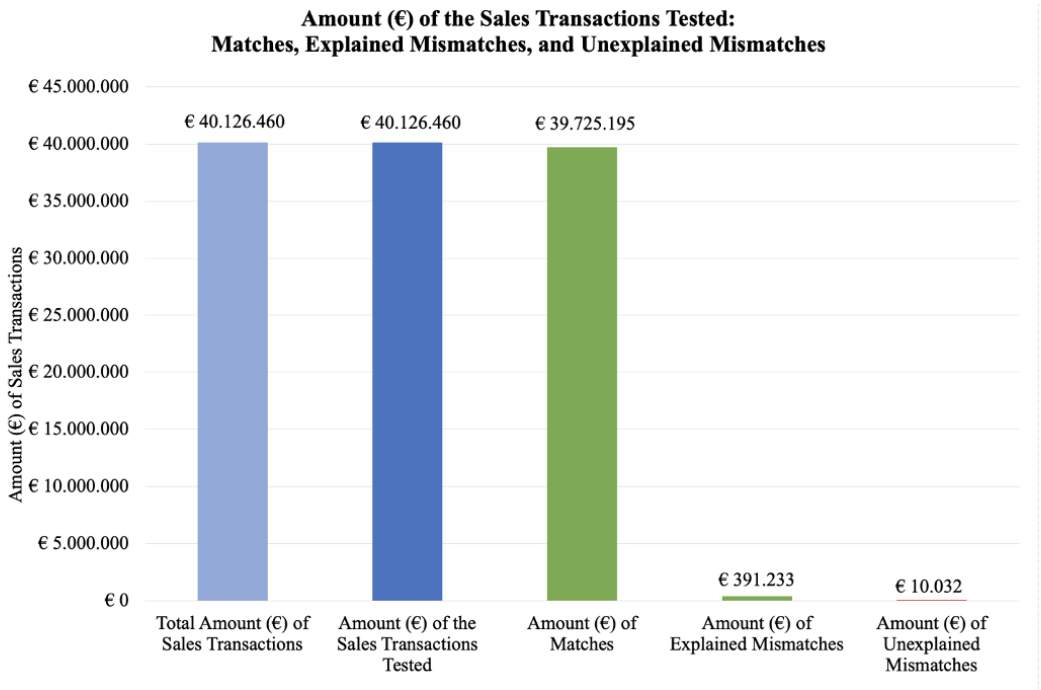
You used a data analytic tool developed by your firm when performing both of the three-way matches on **the entire population** of Ruiter's *Sporting Goods Sales* transactions from 01/01/2022 to 31/10/2022. Each of the three-way matches tested 47.582 sales transactions (the entire population) and the testing covered €40.126.460 (the total amount of the entire population).[15]

B.2 SAMPLE TESTING CONDITION

You performed both of the three-way matches on **a sample** of Ruiter's *Sporting Goods Sales* transactions from 01/01/2022 to 31/10/2022. Each of the three-way matches tested 473 sales transactions (the entire population was 47.582 sales transactions) and the testing covered €401.825 (the total amount of the entire population was €40.126.460).

[15] In our instrument, dates are written in the pattern of day/month/year to be consistent with the date notation format used in the Netherlands. The Dutch numerical format uses periods to separate groups of thousands and uses a comma to indicate the decimal place, opposite to the numerical format used in the U.S.

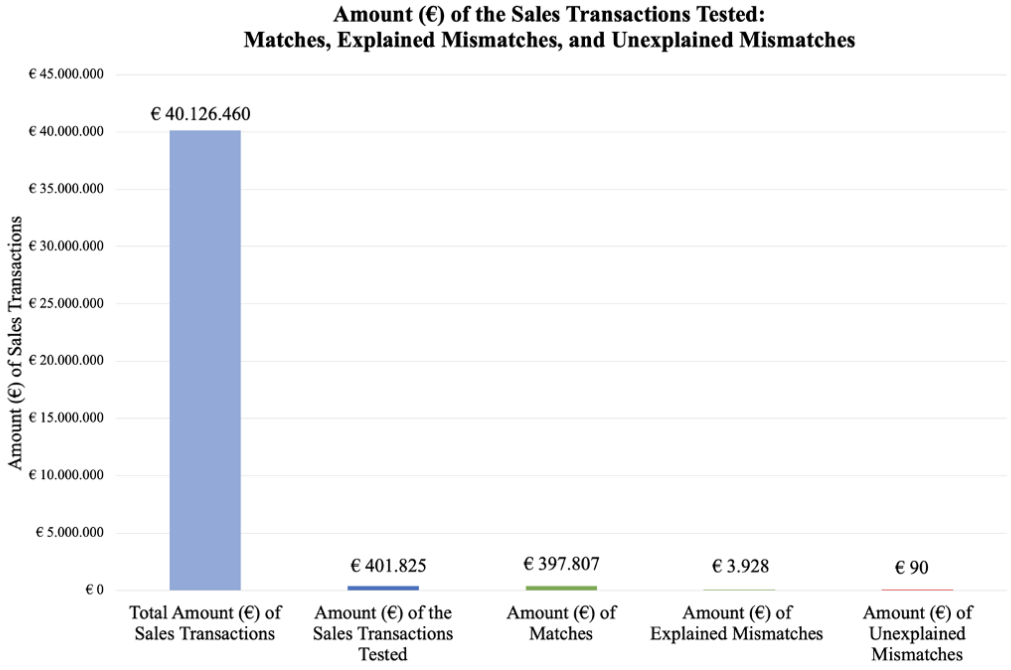
B.3 AN EXAMPLE OF VISUALIZATIONS RECEIVED BY FULL POPULATION TESTING & VISUALIZED RESULTS CONDITION



B.4 AN EXAMPLE OF TABLES RECEIVED BY FULL POPULATION TESTING & TABULATED RESULTS CONDITION

Amount (€) of the Sales Transactions Tested: **Matches, Explained Mismatches, and Unexplained Mismatches**	
Total Amount (€) of Sales Transactions	€ 40.126.460
Amount (€) of the Sales Transactions Tested	€ 40.126.460
Amount (€) of Matches	€ 39.725.195
Amount (€) of Explained Mismatches	€ 391.233
Amount (€) of Unexplained Mismatches	€ 10.032

B.5 AN EXAMPLE OF VISUALIZATIONS RECEIVED BY SAMPLE TESTING & VISUALIZED RESULTS CONDITION



B.6 AN EXAMPLE OF TABLES RECEIVED BY SAMPLE TESTING & TABULATED RESULTS CONDITION

Amount (€) of the Sales Transactions Tested: **Matches, Explained Mismatches, and Unexplained Mismatches**	
Total Amount (€) of Sales Transactions	€ 40.126.460
Amount (€) of the Sales Transactions Tested	€ 401.825
Amount (€) of Matches	€ 397.807
Amount (€) of Explained Mismatches	€ 3.928
Amount (€) of Unexplained Mismatches	€ 90

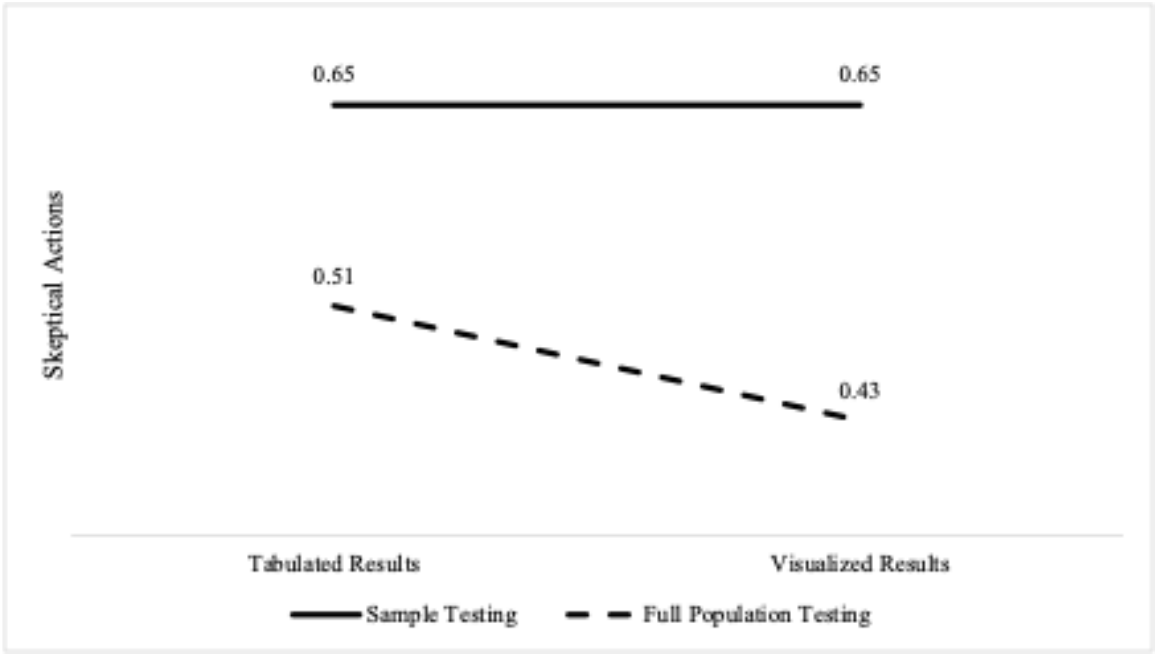
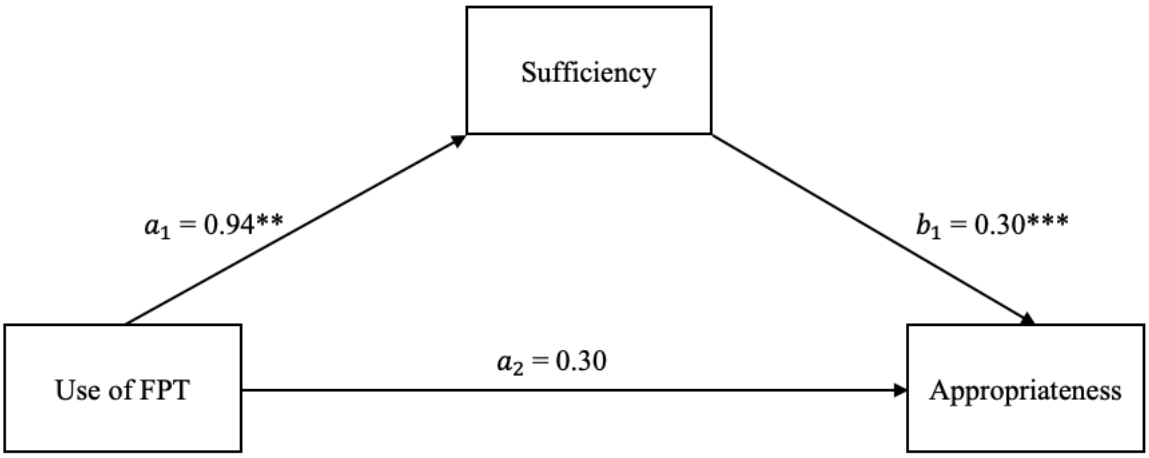


FIG 1. — This figure illustrates the observed pattern of skeptical actions in Full Population Testing & Visualized Results, Full Population Testing & Tabulated Results, Sample Testing & Visualized Results, and Sample Testing & Tabulated Results conditions. *Skeptical Actions* is equal to 1 if participants would inquire of the client management about the industry growth red flag and/or communicate the industry growth red flag to their audit manager, and 0 otherwise.



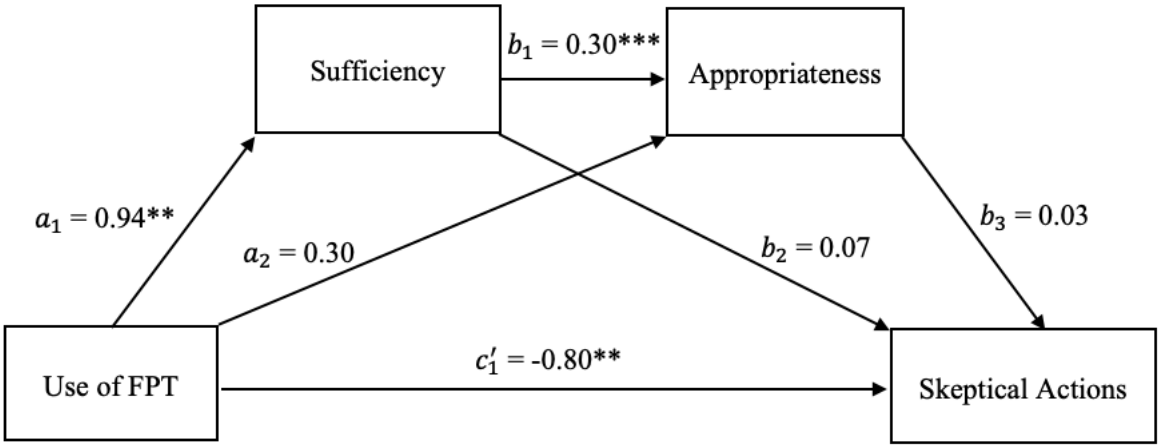
Indirect effect of *Use of FPT* on *Appropriateness* through *Sufficiency* = $a_1 b_1$ = 0.28, Confidence interval = (0.06, 0.61)

FIG 2. — This figure illustrates auditors' substitution of evidence sufficiency evaluation for evidence appropriateness evaluation. *Use of FPT* is equal to 1 if participants were assigned to the *FPT* condition, and 0 for the *Sample Testing* condition. *Sufficiency* is measured by asking participants to what extent they felt that the quantity of sales transactions tested in the two three-way matches at interim testing was sufficient, utilizing an 11-point Likert scale ranging from 0 (Not at all sufficient) to 10 (Completely sufficient). *Appropriateness* is measured by the extent to which participants determined that the evidence obtained from the two three-way match tests was appropriate, utilizing an 11-point Likert scale ranging from 0 (Not at all appropriate) to 10 (Highly appropriate). We used the Hayes (2018) PROCESS model 4 in SPSS and the Preacher and Hayes (2008) bootstrapping approach to test this model. We used 5,000 bootstrap resamples with replacement to estimate 90% confidence intervals. The following equations are used:

$Sufficiency = i_M + a_1 Use\ of\ FPT + \varepsilon$;

$Appropriateness = i_Y + a_2 Use\ of\ FPT + b_1 Sufficiency + \varepsilon$;

***, **, * Indicate significance of coefficients at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively.

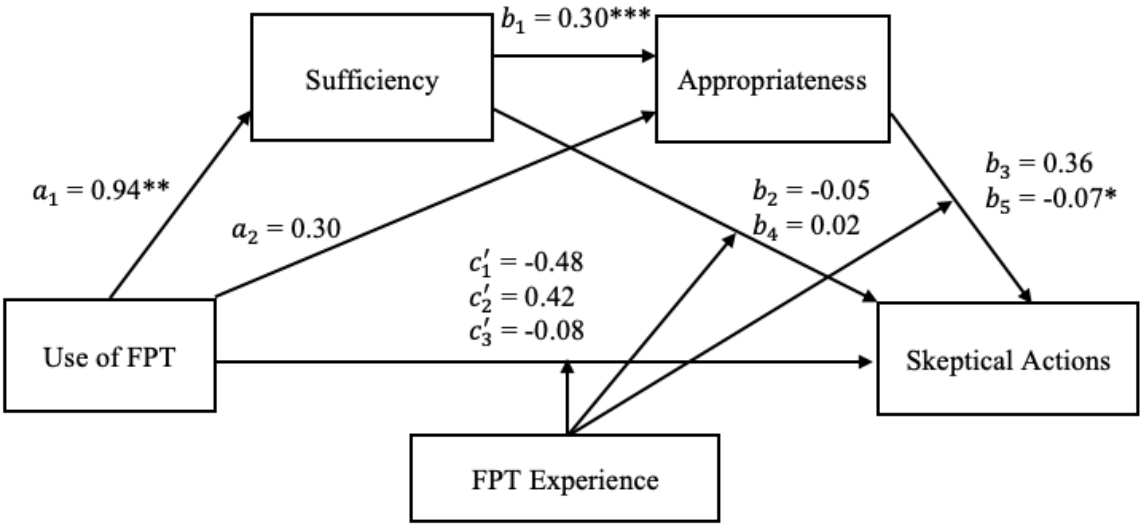


Indirect effect of *Use of FPT* on *Skeptical Actions* through *Sufficiency* = $a_1 b_2$ = 0.07, Confidence interval = (-0.09, 0.25)

Indirect effect of *Use of FPT* on *Skeptical Actions* through *Appropriateness* = $a_2 b_3$ = 0.01 Confidence interval = (-0.05, 0.10)

Indirect effect of *Use of FPT* on *Skeptical Actions* through *Sufficiency* and *Appropriateness* = $a_1 b_1 b_3$ = 0.01, Confidence interval = (-0.05, 0.07)

FIG 3. — This figure illustrates the effect of FPT on skeptical actions through the mediation of evidence sufficiency and appropriateness evaluations in a serial order. *Use of FPT* is equal to 1 if participants were assigned to the *FPT* condition, and 0 for the *Sample Testing* condition. *Sufficiency* is measured by asking participants to what extent they felt that the quantity of sales transactions tested in the two three-way matches at interim testing was sufficient, utilizing an 11-point Likert scale ranging from 0 (Not at all sufficient) to 10 (Completely sufficient). *Appropriateness* is measured by the extent to which participants determined that the evidence obtained from the two three-way match tests was appropriate, utilizing an 11-point Likert scale ranging from 0 (Not at all appropriate) to 10 (Highly appropriate). *Skeptical Actions* is equal to 1 if participants would inquire of the client management about the industry growth red flag and/or communicate the industry growth red flag to their audit manager, and 0 otherwise. We used the Hayes (2018) PROCESS model 6 in SPSS and the Preacher and Hayes (2008) bootstrapping approach to test this model. We used 5,000 bootstrap resamples with replacement to estimate 90% confidence intervals. The following equations are used:
$Sufficiency = i_{M1} + a_1 Use\ of\ FPT + \varepsilon$;
$Appropriateness = i_{M2} + a_2 Use\ of\ FPT + b_1 Sufficiency + \varepsilon$;
$Skeptical\ Actions = i_Y + c_1' Use\ of\ FPT + b_2 Sufficiency + b_3 Appropriateness + \varepsilon$.
***, **, * Indicate significance of coefficients at p < 0.01, p < 0.05, and p < 0.10, respectively.



Index of moderated mediation *FPT Experience* for Path *Use of FPT* → *Sufficiency* → *Skeptical Actions*: Index = 0.02, Confidence interval = (-0.05, 0.11).

Index of moderated mediation *FPT Experience* for Path *Use of FPT* → *Appropriateness* → *Skeptical Actions*: Index = -0.02, Confidence interval = (-0.08, 0.02).

Index of moderated mediation *FPT Experience* for Path *Use of FPT* → *Sufficiency* → *Appropriateness* → *Skeptical Actions*: Index = -0.02, Confidence interval = (-0.06, -0.00).

FIG 4. — This figure illustrates the effect of FPT on skeptical actions through the serial mediation of evidence sufficiency and appropriateness evaluations moderated by participants' experience with using FPT. *Use of FPT* is equal to 1 if participants were assigned to the *FPT* condition, and 0 for the *Sample Testing* condition. *Sufficiency* is measured by asking participants to what extent they felt that the quantity of sales transactions tested in the two three-way matches at interim testing was sufficient, utilizing an 11-point Likert scale ranging from 0 (Not at all sufficient) to 10 (Completely sufficient). *Appropriateness* is measured by the extent to which participants determined that the evidence obtained from the two three-way match tests was appropriate, utilizing an 11-point Likert scale ranging from 0 (Not at all appropriate) to 10 (Highly appropriate). *FPT Experience* captures participants' overall level of experience with performing full population testing, utilizing an 11-point Likert scale ranging from 0 (None) to 10 (Extensive). *Skeptical Actions* is equal to 1 if participants would inquire of the client management about the industry growth red flag and/or communicate the industry growth red flag to their audit manager, and 0 otherwise. We used the Hayes (2018) PROCESS model 89 in SPSS and the Preacher and Hayes (2008) bootstrapping approach to test this model. We used 5,000 bootstrap resamples with replacement to estimate 90% confidence intervals. The following equations are used:
$Sufficiency = i_{M1} + a_1 Use\ of\ FPT + \varepsilon$;
$Appropriateness = i_{M2} + a_2 Use\ of\ FPT + b_1 Sufficiency + \varepsilon$;
$Skeptical\ Actions = i_Y + c_1' Use\ of\ FPT + c_2' FPT\ Experience + c_3' Use\ of\ FPT * FPT\ Experience + b_2 Sufficiency + b_3 Appropriateness + b_4 Sufficiency * FPT\ Experience + b_5 Appropriateness * FPT\ Experience + \varepsilon.$
***, **, * Indicate significance of coefficients at p < 0.01, p < 0.05, and p < 0.10, respectively. The confidence interval related to the index of moderated mediation *FPT Experience* for Path *Use of FPT* → *Sufficiency* → *Appropriateness* → *Skeptical Actions* before rounding is (-0.0599, -0.0005).

TABLE 1

Tests of H1 and H2

Panel A – Descriptive Statistics – *Skeptical Actions*

	Tabulated Results	Visualized Results	Total
FPT	0.51 (0.51) [35]	0.43 (0.50) [30]	0.48 (0.50) [65]
Sample Testing	0.65 (0.49) [34]	0.65 (0.49) [26]	0.65 (0.48) [60]
Total	0.58 (0.50) [69]	0.55 (0.50) [56]	0.56 (0.50) [125]

Panel B – Logistic Regression Results – *Skeptical Actions*

	B	Wald	Df	Sig. (one-tailed)
Use of FPT	-0.73	3.85	1	0.03
Presentation Format	-0.15	0.16	1	0.35
Use of FPT × Presentation Format	-0.36	0.23	1	0.32
Constant	0.26	1.94	1	0.08

This table presents the tests of our H1 and H2. *Use of FPT* is equal to 1 if participants were assigned to the *FPT* condition, and 0 for the *Sample Testing* condition. *Presentation Format* is equal to 1 if the three-way match testing results were visualized, and 0 if tabulated. Panel A presents the cell means, standard deviations, and sample sizes for *Skeptical Actions* in each experimental condition. *Skeptical Actions* is equal to 1 if participants would inquire of the client management about the industry growth red flag and/or communicate the industry growth red flag to their audit manager, and 0 otherwise. Panel B provides the logistic regression results of testing H1, the main effect of *Use of FPT*, and H2, the interaction effect of *Use of FPT* and *Presentation Format*.