An Empirical Study of the Impact of Audit Evidence Reliability on the Accuracy of Audit Opinions

Alana Walsh, Carter Logan, Frederick Lane

1 Introduction

The fundamental purpose of financial statement auditing is to provide reasonable assurance that financial statements are free from material misstatement, whether due to fraud or error. Central to this assurance function is the collection and evaluation of audit evidence, which forms the basis for the auditor's opinion. While auditing standards emphasize the importance of sufficient appropriate audit evidence, the specific relationship between evidence reliability characteristics and the ultimate accuracy of audit opinions remains inadequately explored in empirical literature. This research addresses this critical gap by developing a comprehensive framework for assessing audit evidence reliability and examining its direct impact on audit opinion accuracy.

Traditional auditing research has largely treated audit evidence as a bi-

nary concept—either present or absent—without sufficiently accounting for the nuanced reliability characteristics that distinguish high-quality from low-quality evidence. The auditing profession has long recognized that evidence reliability exists on a continuum, with factors such as source independence, documentation quality, and corroborative strength influencing the weight assigned to different pieces of evidence. However, empirical validation of how these reliability dimensions collectively influence audit outcomes has been limited by methodological constraints and data availability challenges.

This study makes several distinctive contributions to auditing literature. First, we introduce the Audit Evidence Reliability Index (AERI), a novel metric that quantifies evidence quality across multiple dimensions. Second, we employ advanced analytical techniques to model the complex, non-linear relationships between evidence reliability components and audit opinion accuracy. Third, we identify specific threshold effects and interaction patterns that have significant implications for audit efficiency and effectiveness. Fourth, we provide empirical validation of long-standing professional judgments about evidence quality through rigorous statistical analysis.

Our research is situated at the intersection of auditing practice, judgment and decision-making research, and evidence theory. By bridging these domains, we develop a more sophisticated understanding of how evidence characteristics influence auditor judgments and ultimately affect the reliability of financial reporting. The findings have direct relevance for audit standard setters, practicing auditors, audit committees, and financial statement users

who rely on audit opinions for decision-making purposes.

2 Methodology

2.1 Research Design and Data Collection

This study employed a mixed-methods research design, combining quantitative analysis of audit engagement data with qualitative insights from auditor interviews. The primary dataset comprised 347 completed audit engagements from 42 audit firms, representing diverse industries including manufacturing, financial services, technology, and healthcare. Data collection spanned a three-year period from 2020 to 2022, allowing for longitudinal analysis of evidence reliability patterns and their relationship with subsequent audit opinion accuracy.

Engagement selection followed a stratified random sampling approach to ensure representation across firm sizes (Big Four, national firms, and regional firms), industry sectors, and engagement complexity levels. For each engagement, we collected comprehensive documentation including audit programs, working papers, evidence evaluation notes, and final opinion documentation. The accuracy of audit opinions was determined through subsequent events analysis, including restatements, regulatory investigations, and litigation outcomes over a 24-month post-opinion period.

2.2 Audit Evidence Reliability Index (AERI) Development

The cornerstone of our methodological approach was the development of the Audit Evidence Reliability Index (AERI). This composite measure assesses evidence reliability across four primary dimensions: source characteristics, verification methods, temporal relevance, and corroborative relationships. Each dimension was operationalized through multiple indicators, with weights derived from expert judgment surveys and statistical analysis.

The source characteristics dimension evaluated the independence, competence, and objectivity of evidence sources. External confirmation from independent third parties received the highest ratings, while internally generated documentation underwent rigorous assessment of control environment effectiveness. The verification methods dimension assessed the robustness of evidence testing procedures, with direct physical inspection and recomputation scoring higher than analytical procedures alone.

Temporal relevance considered the timing of evidence collection relative to the period under audit, with contemporaneous evidence receiving preferential treatment over retrospective or prospective evidence. The corroborative relationships dimension evaluated how different pieces of evidence reinforced or contradicted each other, with consistent evidence from multiple independent sources achieving the highest reliability scores.

2.3   Analytical Approach

Our analytical strategy employed both traditional econometric models and machine learning techniques to examine the relationship between evidence reliability and audit opinion accuracy. We specified multiple regression models with audit opinion accuracy as the dependent variable and AERI scores as the primary independent variable, controlling for auditor characteristics, engagement complexity, client industry, and temporal effects.

To address potential endogeneity concerns, we implemented instrumental variable approaches using auditor rotation patterns and regulatory inspection cycles as exogenous sources of variation in evidence collection practices. Additionally, we employed propensity score matching to create comparable groups of engagements with high and low evidence reliability characteristics.

The machine learning component utilized gradient boosting algorithms and neural networks to identify complex interaction effects and non-linear relationships that might be missed by traditional linear models. These techniques proved particularly valuable for understanding how different combinations of evidence reliability characteristics influence audit outcomes.

3 Results

3.1 Descriptive Statistics and Reliability Assessment

The analysis revealed substantial variation in evidence reliability across audit engagements, with AERI scores ranging from 0.34 to 0.89 on a normalized 0-1 scale. The distribution exhibited moderate positive skew, indicating that most engagements clustered toward the higher end of the reliability spectrum, though with significant outliers demonstrating exceptionally poor evidence quality.

Reliability across the four AERI dimensions showed interesting patterns. Source characteristics demonstrated the highest average scores, reflecting auditors' general preference for independent evidence sources. Conversely, temporal relevance showed the greatest variability, suggesting inconsistent attention to the timing of evidence collection across different engagements. The internal consistency of the AERI measure was confirmed through Cronbach's alpha coefficients exceeding 0.85 for all primary dimensions.

3.2 Primary Findings: Evidence Reliability and Opinion Accuracy

Our central hypothesis regarding the positive relationship between evidence reliability and audit opinion accuracy received strong support. Regression analysis indicated that a one-standard-deviation increase in AERI scores was

associated with a 27.3

The analysis revealed important non-linearities in this relationship. Specifically, we identified diminishing returns to evidence reliability beyond certain thresholds. Engagements with AERI scores above 0.75 showed minimal additional improvements in opinion accuracy, suggesting the existence of a reliability sufficiency threshold. Conversely, engagements scoring below 0.45 demonstrated dramatically higher rates of opinion inaccuracy, indicating a critical minimum reliability level necessary for effective auditing.

3.3   Dimensional Analysis and Interaction Effects

Examination of individual AERI dimensions revealed differential impacts on opinion accuracy. Source characteristics emerged as the strongest predictor, with independent external evidence contributing disproportionately to overall reliability. The verification methods dimension showed interesting interaction effects with engagement complexity, becoming increasingly important in more complex audit environments.

A particularly novel finding concerned what we term 'evidence reliability cascades'—situations where high-reliability evidence in one area positively influenced the perceived reliability of adjacent evidence. For example, strong external confirmation of account balances appeared to enhance auditors' assessment of internally generated documentation reliability, even when the connection between evidence types was somewhat tenuous.

3.4 Machine Learning Insights

The gradient boosting analysis identified several complex interaction patterns that traditional models failed to capture. Most notably, the relationship between evidence quantity and reliability demonstrated clear threshold effects, with additional evidence collection yielding minimal benefits once certain reliability thresholds were achieved. This finding challenges conventional audit approaches that emphasize comprehensive evidence collection without sufficient attention to reliability optimization.

The neural network models successfully predicted audit opinion accuracy with 89.7

4 Conclusion

This research makes several significant contributions to auditing theory and practice. By developing and validating the Audit Evidence Reliability Index, we provide auditors with a practical tool for assessing and improving evidence quality throughout the engagement. The empirical demonstration of the strong relationship between evidence reliability and audit opinion accuracy reinforces professional standards while providing quantitative guidance for evidence collection decisions.

The identification of reliability thresholds and diminishing returns has important implications for audit efficiency. Rather than pursuing exhaustive evidence collection, auditors may achieve better outcomes by focusing

on high-reliability evidence sources once minimum sufficiency thresholds are met. This insight could significantly enhance audit productivity without compromising quality.

The concept of evidence reliability cascades introduces a new dimension to evidence evaluation, suggesting that strategic placement of high-reliability evidence may positively influence the entire audit process. This finding aligns with cognitive psychology research on anchoring and adjustment heuristics, providing a theoretical foundation for observed auditor judgment patterns.

Several limitations warrant consideration. The sample, while diverse, may not fully represent all audit environments, particularly in emerging markets or highly specialized industries. The retrospective nature of opinion accuracy assessment introduces potential measurement error, though we employed multiple validation approaches to mitigate this concern. Future research could extend our framework to different cultural contexts, explore dynamic evidence reliability assessment throughout the audit process, and examine how technological advancements in data analytics influence evidence reliability perceptions.

In conclusion, this study establishes that audit evidence reliability systematically influences audit opinion accuracy in predictable and measurable ways. By quantifying this relationship and identifying specific mechanisms through which reliability affects outcomes, we provide both theoretical insights and practical tools for enhancing audit quality. As the auditing profession continues to evolve in response to technological change and regulatory

pressures, attention to evidence reliability fundamentals remains essential for maintaining public trust in financial reporting.

References

Ahmad, H. S., Zafar, R., Tariq, H. (2017). Fraud detection through continuous auditing and monitoring in the banking sector. Journal of Financial Crime, 24(4), 672-687.

Allen, R. D., Hermanson, D. R. (2021). The evolving role of data analytics in external auditing. Journal of Information Systems, 35(2), 1-15.

Brazel, J. F., Agoglia, C. P. (2022). The effects of audit evidence reliability and complexity on auditor judgments. Auditing: A Journal of Practice Theory, 41(1), 23-45.

Carpenter, T. D., Reimers, J. L. (2019). Professional skepticism and audit evidence evaluation. Accounting Horizons, 33(2), 15-32.

Griffith, E. E., Hammersley, J. S., Kadous, K. (2020). The effects of evidence reliability and subjectivity on auditor evaluations of evidence. Contemporary Accounting Research, 37(3), 1567-1595.

Knechel, W. R., Salterio, S. E. (2021). Auditing: Assurance and risk (5th ed.). Routledge.

Messier, W. F., Glover, S. M., Prawitt, D. F. (2022). Auditing and assurance services: A systematic approach (11th ed.). McGraw-Hill Education.

Peecher, M. E., Schwartz, R., Solomon, I. (2019). It's all about audit

quality: Perspectives from strategic audits. Perspectives on Auditing, 3(1), 1-25.

Trotman, K. T., Bauer, T. D., Humphreys, K. A. (2020). Group judgment and decision making in auditing: Research and practice. Auditing: A Journal of Practice Theory, 39(3), 45-68.

Zimbelman, M. F., Waller, W. S. (2021). Auditing and accounting education: Developing professional judgment. Issues in Accounting Education, 36(2), 1-18.