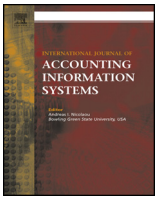
International Journal of Accounting Information Systems 24 (2017) 1–14

Contents lists available at ScienceDirect

International Journal of Accounting Information Systems

journal homepage: www.elsevier.com/locate/accinf



External auditors' evaluation of the internal audit function: An empirical investigation

Renu Desai [a], Vikram Desai [a,*], Theresa Libby [b], Rajendra P. Srivastava [c]

[a] Nova Southeastern University, 3301 College Avenue, Davie 33314, United States
[b] University of Waterloo, Waterloo, ON N2L 3G1, Canada
[c] University of Kansas, 1300 Sunnyside Av, Lawrence, KS 66045, United States

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 10 December 2016
Accepted 18 December 2016
Available online xxxx

JEL classification:
M42

Keywords:
Internal audit function
Reliance on work of the internal auditor
Belief functions

ABSTRACT

Evaluation of the strength of the client's internal audit function by the external auditor (EA) has taken on increased significance due to stronger regulation around the evaluation of internal controls after SOX (2002). However, research examining how this evaluation occurs in practice is mixed and inconclusive. In this study, we examine empirically whether the Desai et al. (2010) theoretical model is reflective of how auditors make judgments about the strength of their client's internal audit function in practice. Specifically, we present external auditors with evidence about internal auditor work performance, competence and objectivity in a manner consistent with the structure of evidence evaluation implied by the Desai et al. (2010) model. We then compare the auditors' actual strength judgments to the strength levels predicted by the model and evaluate similarities and differences. Results indicate that no one factor dominates the strength judgment in all cases. In addition, EAs do not weigh negative evidence as heavily as does the model. When the evidence about the three factors is conflicting, external auditors have difficulty incorporating them in a consistent way into the calculation of their overall strength judgment. Finally, we find results consistent with prior research indicating auditors tend to be more sensitive to negative than positive evidence. Also, it is harder to move auditors' beliefs away from a negative position with positive evidence than to move those beliefs away from a positive position with negative evidence. Results suggest that additional training and use of a decision aid structured according to the Desai et al. (2010) model would be especially useful when evidence about internal auditors' work performance, competence and objectivity is conflicting.

1. Introduction

Evaluation of the strength of the internal audit function (IAF) has taken on increased significance due to stronger regulation around the evaluation of internal controls after SOX (2002) (Desai et al., 2011). When the IAF is strong, external auditors may choose to rely on their work (Public Company Accounting Oversight Board, PCAOB, 2007; International Federation of Accountants, IFAC, 2012). Recently, the literature has highlighted the need for more research on the degree of reliance the external auditor places on the work of the internal auditor (Messier et al., 2011; Weisner and Sutton, 2015). In addition, results of

* Corresponding author.
E-mail addresses: rdesai@nova.edu (R. Desai), vdesai@nova.edu (V. Desai), talibby@uwaterloo.ca (T. Libby), rsrivastava@ku.edu (R.P. Srivastava).

http://dx.doi.org/10.1016/j.accinf.2016.12.001
1467-0895/

practice inspection reports of the Public Company Accounting Oversight Board (PCAOB) indicate external auditors often rely either too much or too little on the work of the internal auditor (Public Company Accounting Oversight Board, PCAOB, 2009). Understanding the structure of the external auditor's strength judgment is a crucial first step to understanding their ultimate IAF reliance decision (Gramling et al., 2004)

Bame-Aldred et al. (2012) note that results of prior studies examining how external auditors judge the strength of their clients' IAF have been mixed and inconclusive. According to Krishnamoorthy (2002), this is likely because the interrelationships between the three internal auditor qualities that auditing standards dictate must be assessed by external auditors, i.e., work performance (W), competence (C) and objectivity (O) are complex. In addition, the way external auditors make judgments of the strength of the client's IAF (i.e., how evidence is combined, weighted etc.) is largely unobservable. Thus, it is difficult to break apart the final strength judgment into component parts without some guidance from theory. Desai et al. (2010) develop a prescriptive model based on belief function theory [see Shafer (1976) and Srivastava and Mock (2002)] to better understand how auditors make these complex judgments. They argue that this model, due to its conceptual and intuitive appeal, is a particularly good candidate to act as the basis for an audit decision aid designed to guide auditors in forming this relatively complex, multi-input judgment.[1]

The objective of the current study is to examine how the Desai et al. (2010) model performs relative to the judgments of a group of external auditors experienced in evaluating the strength of their clients' IAFs. We focus on two specific areas where we believe the strength outputs of the model may differ from the judgments of our sample of experienced external auditors. First, the model assumes that the impact of positive and negative evidence about the quality of the IAF on auditors' judgments regarding the inter-relationships between W, C and O will be symmetric.[2] Since auditors are trained to be skeptical, it is likely they will react more strongly to negative evidence than they will to positive evidence of the same magnitude (Butt and Campbell, 1989; Ashton and Ashton, 1988; Knechel and Messier, 1990). We formally test this prediction in H1.

Second, the IAF strength judgment is complex and involves the evaluation and weighting of several preliminary judgments into one overall judgment. When individuals are faced with complex decisions, they tend to fall back on heuristics or simplifying decision rules (Payne et al., 1993). This can reduce the quality of decisions if individuals don't take into account all available pieces of information. In the current context, IAF strength judgments based on convergent judgments about W, C and O (i.e., W, C and O are all judged to be high or low) are likely to be much less complex than are judgments based on conflicting evidence (i.e., some of W, C and O are judged to be high while others are low). Therefore, we predict that the judgments of our external auditor participants will be significantly different from the beliefs derived from the model when judgments about W, C and O are conflicting. We formally test this prediction in H2.

To test our hypotheses, we construct a case scenario based on SAS No. 65 (American Institute of Certified Public Accountants, AICPA, 1991) and prior research (Brown, 1983; Schneider, 1984, 1985; Margheim, 1986; Messier and Schneider, 1988; Maletta, 1993) in which we manipulate W, C and O in a 2 (work performance high/low) × 2 (competence high/low) × 2 (objectivity high/low) between subjects design.[3] A total of 109 US-based external auditors, all of whom are experienced in making judgments about the strength of their clients' IAF, took part in the experiment. The participants evaluate positive and/or negative evidence about W, C and O depending on the experimental condition to which they are randomly assigned and then judge the strength of the hypothetical client's IAF.

The experimental results yield several interesting findings. First, we find some support for H1 in that the interrelationships between W and C (r_1) and between W and O (r_2) are asymmetric in their influence on the external auditors' beliefs. These results are consistent with prior research indicating auditors tend to be more sensitive to negative than positive evidence (Ashton and Ashton, 1988; Butt and Campbell, 1989). In addition, these results are consistent with prior research indicating it is harder to move auditors' beliefs away from a negative position with positive evidence than to move those beliefs away from a positive position with negative evidence (Ashton and Ashton, 1988; Asare, 1992; McMillan and White, 1993; Knechel and Messier, 1990).

Second, we compare the strength judgments of our auditor participants to the related strength outputs derived from the model. As predicted in H2, when judgments about W, C and O are conflicting (i.e., some positive and some negative), participants recognize the interrelationships between the factors, but appear to have difficulty incorporating them in a manner consistent with the model into the calculation of their overall strength judgment, likely due to cognitive limitations (Payne et al., 1993). We also observe that when judgments about W, C and O are all positive, the observed and derived beliefs are not significantly different from one another. On the other hand, when judgments about W, C and O are all negative, our auditor participants do not weigh negative evidence as heavily as does the model. These results taken together suggest that auditors in practice might, on average, develop less negative beliefs in the weakness of the IA function than the model suggests they should. Thus, they many choose to rely on the IA's work even when it is unwarranted. These results suggest that a decision aid would be a useful tool to improve these complex judgments, especially when evidence about W, C and O is conflicting.

This paper contributes to both research and practice. From the perspective of research, we are unaware of other studies that have attempted to empirically estimate the impact of evidence about the quality of the IAF on the interrelationships between the

[1] We interpret the nature of the decision aid implied by Desai et al. (2010) to be similar to a decision support system as defined by Rose (2002) which provides a meaningful structure for applying judgment in an otherwise highly unstructured decision context.

[2] We note judgments about the inter-relationships between W, C and O are normally unobservable in practice, but the structural definitions of these inter-relationships (i.e., r_1 and r_2) as laid out by Desai et al. (2010) allow us to estimate them in our experimental setting, as described in the methods section (Section 3).

[3] SAS No. 128 (American Institute of Certified Public Accountants, AICPA, 2014) supersedes SAS No. 65, but is effective for year ends on or after December 15, 2014. The key difference between the standards is that SAS 128 requires the external auditor to evaluate whether the IAF is applying a systematic and disciplined approach to planning, performing, supervising, reviewing and documenting its activities, including quality control. We do not ask our external auditor participants to make this judgment, but given our data were collected before the effective date of the new standard, we rely on SAS No. 65 for guidance.

three factors W, C, and O. Neither has prior research studied the effect of convergent and conflicting evidence regarding W, C and O on the external auditors' evaluation of the strength of the IA function. As noted by Bame-Aldred et al. (2012), "…we continue to know very little about how and to what extent EAs are currently evaluating IAF quality factors. (266)." In addition, Harrison et al. (2002) argue the question of whether auditors combine evidence consistent with the axioms of the belief function framework is an important and open question requiring additional research. We provide some of the first empirical evidence of the way auditors judge the strength of the interrelationships between W and C and between W and O based on the nature (i.e., convergent or conflicting) of the underlying pieces of evidence. These strength judgments, along with materiality and account-level risk, then act as inputs into the external auditor's decision to rely on the IAF's work. Our results are important given auditing standards do not clearly indicate how intermediate judgments of the quality of W, C and O should interrelate when evaluating the strength of the IA function and results from previous studies of this issue have been mixed and inconclusive (Gramling et al., 2004, Krishnamoorthy, 2002).

By providing initial empirical evidence on (in)consistencies between auditors' judgments and model predictions, we move the literature one step closer to the development of a useful decision aid that would be appealing to practitioners who wish to test the validity of their judgments of IAF strength. Further, the use of decision aids to enhance audit effectiveness has been increasing in the post-SOX reporting climate (see for example, Bell and Carcello, 2000; Bell et al., 2002).[4] Our results suggest that auditors in the field could benefit from a model-based decision aid to guide their evaluation of the various indicators of IAF, especially in the case of conflicting evidence. In addition, such a decision aid could be extremely useful to novice auditors as a means to test their initial IA strength judgments as they work to develop their skills and judgment through training and experience (Rose et al., 2007; Arnold et al., 2004). Developing a better understanding of auditor judgments about the strength of the IAF relative to the beliefs derived from the model is an important first step in the design of such a decision aid.

The rest of the paper is organized as follows. The next section reviews the literature and the development of our hypotheses. Section 3 describes the methodology and experimental design and presents the results. The paper concludes with a discussion of the results and their implications for research and practice.

2. Literature review and hypothesis development

2.1. Structure of the Desai et al. (2010) model

Desai et al. (2010) summarize the literature identifying the pieces of evidence typically gathered by the external auditor when evaluating the internal auditors' work performance (W), objectivity (O), and competence (C). They model the structure of the interrelationships between the three factors using the evidential reasoning approach of Srivastava et al. (2007) under the Dempster-Shafer theory of belief functions.[5] The model assumes the evidence collected by the external auditor (i.e., the rectangular boxes in Fig. 1) will indicate one of two values for W, C and O: high (H) or low (L). Based on the evidence gathered, the internal auditor's work performance is evaluated to be satisfactory or unsatisfactory (W_H or W_L), they are evaluated to be either competent or incompetent (C_H or C_L), and they may be evaluated as objective or lacking in objectivity (O_H or O_L).

The external auditor collects evidence and uses it to develop beliefs about W $(m_w^+, m_w^-, m_w^\Theta)$, C$(m_c^+, m_c^-, m_c^\Theta)$, and O$(m_o^+, m_o^-, m_o^\Theta)$ where + indicates a belief that the factor is high or strong, − indicates a belief that the factor is low or weak, and Θ represents lack of knowledge (also called "ignorance" in the belief function literature). To illustrate, suppose the external auditor gathers evidence about the competence of the internal auditor and develops a belief that the internal auditor is competent (C_H) with a confidence level of 0.6, on a scale of 0 to 1. Moreover, the external auditor thinks that the evidence does not provide any support (belief = 0) in favour of the internal auditor being incompetent (C_L). Under belief functions, we can write the above judgment in terms of belief masses as: $m(C_H) = 0.6$, and $m(C_L) = 0$. The unassigned mass of 0.4 is associated with the entire frame $\{C_H, C_L\}$ and represents ignorance (i.e. no information). There are many real-world situations including auditing where such partial ignorance or ambiguity is common and the conventional probability framework is not appropriate to model such uncertainties (see, e.g., Srivastava and Mock, 2002). These beliefs are important inputs to the external auditor's overall evaluation of the strength of the IA function.

According to Desai et al. (2010), the external auditor's belief about the strength of the IA function can be strong, Bel (S_S) or weak, Bel (S_W), depending upon the evidence gathered about work performance, competence and objectivity, and upon the strength of the perceived interrelationships (r_1 and r_2) between them. The external auditor's overall belief about the strength of the IA function can then be predicted as follow:

$$\mathrm{Bel}(S_S) = \left[m_C^+ m_W^+ m_O^+ + r_1 m_C^\Theta m_W^+ m_O^+ + r_2 m_C^+ m_W^+ m_O^\Theta + (r_1 + r_2 - r_1 r_2) m_C^+ m_W^\Theta m_O^+ + r_1 r_2 \left(m_C^+ m_W^\Theta m_O^\Theta + m_C^\Theta m_W^+ m_O^\Theta + m_C^\Theta m_W^\Theta m_O^+\right)\right] / K \quad (1)$$

[4] The PCAOB also indicates the reliance decision should be risk-based. The case-based scenario presented to participants in our experiment was designed such that risk was low and thus, standards suggest the external auditor can rely on the work of the IAF if their evaluation of the IAF is strong.

[5] We incorporate evidence consistent with that listed in Desai et al. (2010) except that related to the evidence concerning corporate governance and its relation to the quality of W and O. This link is less well established in the literature and requires additional information be included in the case scenario regarding organizational structure and reporting relationships making the case scenario more complex and adding noise to the results. Thus, we leave tests related to this type of evidence to future research.

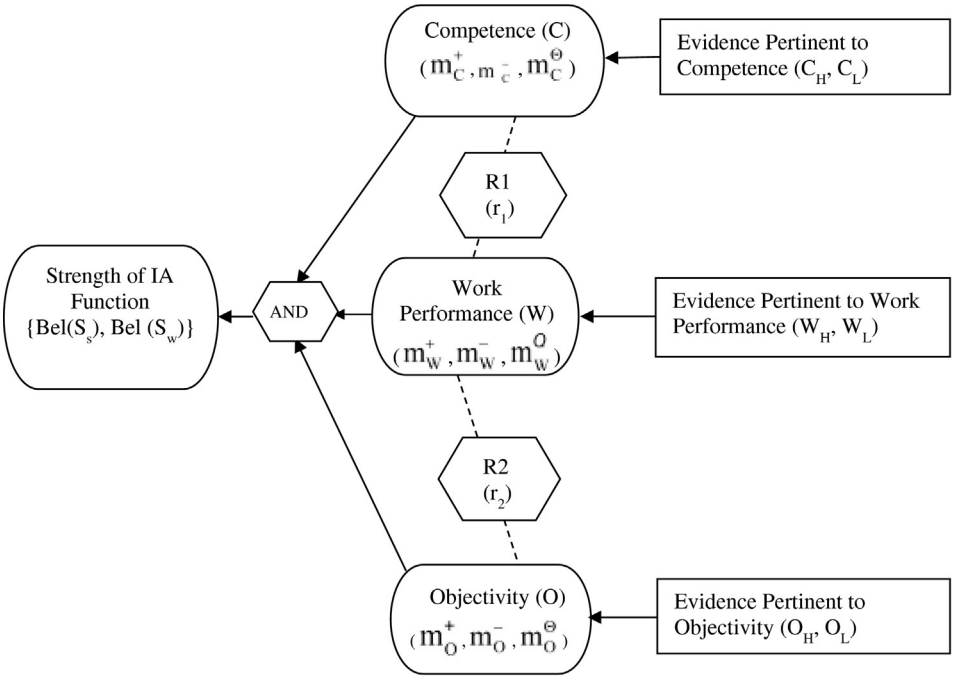


Fig. 1. Desai et al. (2010) Conceptual model Definitions: (C_H, C_L) Evidence that internal auditors are competent (C_H) or incompetent (C_L) (W_H, W_L) Evidence that internal auditors' work performance is satisfactory (W_H) or unsatisfactory (W_L) (O_H, O_L) Evidence that internal auditors are objective (O_H) or not objective (O_L) (m_C^+,m_C^-,m_C^Θ) The belief that the internal auditor is competent (m_C^+), incompetent (m_C^-), or "ignorance" about competence (m_C^Θ) (m_W^+,m_W^-,m_W^Θ) The belief that the internal auditor's work performance is satisfactory (m_W^+) or unsatisfactory (m_W^-) or "ignorance" about work performance (m_W^Θ) (m_O^+,m_O^-,m_O^Θ) The belief that the internal auditor is objective (m_O^+) or lacking in objectivity (m_O^-) or "ignorance" about objectivity (m_O^Θ) r_1 The strength of the relationship between belief about work performance and belief about competence r_2 The strength of the relationship between belief about work performance and belief about objectivity {Bel(S_S), Bel(S_W)} Belief in the overall strength of IA function being strong (S_S) or being weak (S_W). AND Structure of the formulation of the external auditor's overall belief in the strength of the IA function based on the evidence collected. IA function will be evaluated as strong *iff* evidence indicates internal auditors are competent, objective and perform work of satisfactory quality (per set notation $S_S = C_H \cap W_H \cap O_L$).

$$\mathrm{Bel}(S_W) = \left(m_C^+ + m_C^\Theta\right)\left(m_W^+ + m_W^\Theta\right)\left(m_O^+ + m_O^\Theta\right)/K \tag{2}$$

where $\{m_C^+, m_C^-, m_C^\Theta\}$, $\{m_W^+, m_W^-, m_W^\Theta\}$, and $\{m_O^+, m_O^-, m_O^\Theta\}$, respectively, represent the input belief masses related to W, C and O. The symbol K is the renormalization constant and is defined as:

$$\begin{aligned} K = {} & r_1 m_O^\Theta\left(m_C^+ m_W^- + m_C^- m_W^+\right) - r_2 m_C^\Theta\left(m_W^+ m_O^- + m_W^- m_O^+\right) \\ & r_1 r_2 m_W^\Theta\left(m_C^+ m_O^- + m_C^- m_O^+\right) - r_1\left(m_C^- m_W^+ m_O^+ + m_C^+ m_W^- m_O^-\right) \\ & r_2\left(m_C^+ m_W^+ m_O^- + m_C^- m_W^- m_O^+\right) - (r_1 + r_2 - r_1 r_2)\left(m_C^+ m_W^- m_O^- + m_C^- m_W^+ m_O^-\right) \end{aligned} \tag{3}$$

The first term in Eq. (1) arises directly because of the logical "And" relationship between the IA function's strength (S) and the three factors W, C and O. The next three terms in Eq. (1) are included because, due to interrelationships among factors, even if we do not know about one factor we can infer that the factor is present based on the presence of the other two factors. The last term in Eq. (1) represents the situation where we do not have information about two of the factors, but because of the interaction, the knowledge of the presence of one factor can provide knowledge about the presence of the other two factors.

Based on the evidence collected, it is possible for the external auditor to form a belief about the strength of r_1, the interrelationship between W and C, and r_2, the interrelationship between W and O. According to Desai et al. (2010) and tenets of the belief function formalism, the values of r_1 and r_2 will range from zero to one. A value of zero implies there is no relationship and a value of one implies there is a perfect relationship between the two factors. Although Maletta (1993) and Krishnamoorthy (2002) explicitly recognize the interrelationships between factors, prior research does not attempt to understand the interrelationships and how the interactions between them can help auditors gain an understanding of the internal control structure of the client (Brown, 1983; Abdel-Khalik et al., 1983; Schneider, 1984, 1985; Margheim, 1986; Messier and Schneider, 1988; Edge and Farley, 1991). According to Desai et al. (2010), the lack of exploration of these interrelationships in prior research has resulted in mixed and inconsistent findings about external auditor evaluation of the IAF.

2.2. Hypothesis development

The validity of the Desai et al. (2010) model as a basis for a decision aid for auditors in the field is dependent upon the consistency of the model's structure with actual experienced external auditor judgment. There are several indications from the prior literature that a model based on the belief function formalism would be a good fit to actual auditor judgment. First, evidence in the psychology literature (Curley, 2007; Curley and Golden, 1994) suggests that decision makers tend to think in terms of belief functions rather than in Bayesian probability terms. This implies that the belief function representation of the problem of making judgment about the strength of the IAF based on evidence collected about W, C and O, may be more consistent with the auditor's own mental representation of the problem than a Bayesian formulation.

Second, belief function theory allows for judgments about the strength of the evidence collected on a scale between zero and one, a format that is likely more familiar to auditors than the conditional probability judgments required under Bayesian analysis (Srivastava and Mock, 2000).[6] Krishnamoorthy (2002) argues that conditional probability judgments required for Bayesian analysis are not typically available to the auditor when making judgments of the strength of the IAF. Third, Fukukawa and Mock (2011) find that auditors forming belief assessments based on underlying evidence are better able to assess the direction of the evidence (i.e., confirming or disconfirming) than auditors forming Bayesian assessments of risk. Finally, Harrison et al. (2002) find that >80% of their auditor participants' judgments about evaluation of audit evidence were consistent with the belief function rather than the Bayesian framework.

There are some areas, however, where the model may not be consistent with actual auditor beliefs. First, Desai et al. (2010) model each of the interrelationships r_1 and r_2 to be symmetric in their influence on the external auditor's overall strength judgment. In other words, positive evidence and negative evidence of equal strength should have a similar effect on the magnitude of r_1 and r_2. While this assumption aids in the tractability of their mathematical model, this assumption is not without tension given that auditors are trained to be skeptical (Hurtt et al., 2013). Skepticism may make auditors particularly sensitive to evidence that is negative with respect to the hypothesis that financial statements are fairly presented or that controls are adequate, since the primary legal and professional risk faced by auditors is that errors or irregularities in financial statements will go undetected. It follows therefore that external auditors will react more strongly to negative than positive evidence of similar magnitude (Asare, 1992; Ashton and Ashton, 1988; Knechel and Messier, 1990; McMillan and White, 1993) which would be inconsistent with the assumption of Desai et al. (2010). Therefore, we hypothesize the following:

H1. When evidence about the internal auditor's work is negative, external auditors will assign higher values to r_1 and r_2 than when evidence about the internal auditor's work is positive.

An additional reason why the model may generate different indicators of IAF strength than will experienced auditors is that the cognitive complexity of combining, weighting and analyzing multiple pieces of evidence into one overall judgment is cognitive complexity. Dawes (1979) states that the limited ability of people to integrate information from different sources and arrive at decisions is the most consistent finding in prior psychology research. When individuals are faced with complex decisions, they tend to fall back on heuristics or simplifying decision rules (Payne et al., 1993), such as satisficing (Simon, 1957), equal weighting of all attributes (Einhorn and Hogarth, 1975) or "follow the majority of confirming dimensions" (Russo and Dosher, 1983).[7] Consistent with psychology research, Libby (1981) observed that accountants are much better at coding and gathering evidence than integrating it to make optimal decisions when the evidence is conflicting. In the current context, we argue that IAF strength (weakness) judgments which have as inputs convergent positive (negative) evidence about W, C and O (i.e., all high or all low) will be much less complex than when inputs are conflicting (i.e., some of W, C and O are high while others are low) (Asare and Wright, 1997). While the external auditors may be adept at gathering both positive and negative evidence about W, C and O, they may not be as good at integrating this evidence to arrive at a good quality judgment about the strength of the IAF, especially when the evidence is conflicting in nature. Therefore, we predict that the judgments of our external auditor participants will be more similar to the output of the model when evidence about W, C and O are convergent than when evidence about W, C and O is conflicting. These arguments lead to the following predictions.

H2. When evidence about the internal auditor's work performance, competence and/or objectivity is conflicting, the external auditor's overall belief in the strength of the IAF will be significantly different from the belief derived using the Desai et al. (2010) model. 3.

3. Method

3.1. Design and pretest

We conduct a 2 (work performance low/high) × 2 (competence low/high) × 2 (objectivity low/high) between-subjects experiment.[8] Participants read a case scenario describing Barston Inc. as follows:

[6] Specifically, the framework provides "… a more flexible and adaptable way to combine evidence from a variety of sources" (Srivastava and Mock, 2000, 226), allows a better method for mapping uncertainty in accounting and auditing judgments (Harrison et al., 2002) and allows for explicit assessments of ambiguity which are implicit, but not easily accessed from Bayesian assessments (Fukukawa and Mock, 2011; Srivastava, 1993).

[7] Each of these heuristics is described in detail in Payne et al. (1993).

[8] While Margheim (1986) argues that competence and work performance should be either both high or both low since other combinations are rare (e.g. low competence but high work performance), we chose a fully crossed design as a more complete test of our hypotheses.

Barston Inc. is a large public company that manufactures and distributes medical equipment to hospitals, nursing homes and large medical practices. The company has been in business since 1979 and has been publicly traded since 1990. Although Barston Inc. has several large competitors, the company has been able to maintain a 40% market share in the US market throughout the last 15 years.

Participants act in the role of the external auditor involved in planning the current year's audit of Barston Inc. Participants are provided with pieces of evidence about the IAF work performance, competence and objectivity and are asked to express their beliefs about the overall strength of Barston's IAF.[9]

Desai et al. (2010) define the various pieces of evidence that are indicative of the quality of the IAF W, C and O consistent with SAS No. 65 (AICPA 1991) (as in Fig. 1).[10] Based on their work and prior experimental research on this topic (Maletta, 1993; Messier and Schneider, 1988; Clark et al., 1980), we manipulate evidence of high or low quality work performance as a function of internal auditor effort (e.g., percent of time performing planned audit work), execution of the internal audit plan (e.g., completion of audit plan each year versus not) and thoroughness and quality of internal audit reporting (e.g., detailed examination and use of appropriate sampling and EDP audit techniques and detailed documentation versus a general overview and observation with little documentation completed).[11]

The manipulation of high or low IAF competence is based on Desai et al. (2010) and prior work in this area (Gramling et al., 1997; Maletta, 1993; Messier and Schneider, 1988; Brown, 1983). It includes evidence of auditor qualification and training (e.g., mainly former Big 4 auditors with at least 3 years of Big 4 experience and most hold CIA certification versus recent college graduates with degrees in accounting but no work experience), audit planning and supervision (e.g., experienced, CIA qualified department head who plans and supervises work versus inexperienced, not CIA qualified department head who waits for departmental requests before performing audit work) and auditor tacit knowledge (e.g., in-house training versus no in-house training provided). Finally, based on Dezoort et al. (2001) and Maletta (1993), we manipulate high and low levels of IA objectivity as a function of the managerial reporting relationship (e.g., IAF reports to the audit committee versus the company CFO), the breadth and depth of the investigatory scope (e.g. IAF defines plan versus CFO reviews and approves plan) and level of follow-up on recommendations (e.g., audit committee responsible for follow-up versus senior management group including CFO responsible for follow-up).[12]

We performed two types of pretests of our experimental instrument. The first was conducted with a group of 60 senior undergraduate auditing students who evaluated whether the pieces of evidence about IAF quality (i.e., W, P and O) that we included in our experimental instrument were positive or negative. We found that the positive evidence was judged as positive by these participants and the negative evidence was judged as negative. The second type of pretest was conducted with five faculty members with significant experience in teaching the subject of auditing. They agreed that the set of evidence provided in our manipulations were relevant and typical of the type of evidence gathered in practice. In addition, they judged the descriptions of positive and negative evidence to be parallel in nature, that is, the words used to describe the positive and negative aspects of the evidence in each of the manipulations did not differ in strength or implied severity across the positive and negative versions of our manipulations. A few minor adjustments were made to the final instrument based on the results of the pre-tests and it was then uploaded to a project-specific web site. Another group of ten accounting faculty and PhD students then completed the experiment on-line to test the reliability and functionality of the web site.

3.2. Ordering the evidence

We formulated our research question to specifically address how the external auditor would evaluate the strength of their client's IAF when deciding whether to rely on the IAs work. Thus, we needed to decide in which order to present our experimental participants with evidence about W, C and O. We note that auditing standards do not require auditors to collect evidence about W, C and O in any particular order. Even so, we are interested in evaluating the otherwise unobservable interrelationships r_1 and r_2 indicated by the Desai et al. (2010) model. It is our view that presenting evidence about work performance first facilitates this evaluation.

We reinforced our confidence that collecting evidence about work performance first does not pose a significant threat to the validity of our results in two ways. First, we conducted interviews with 5 Big 4 audit partners who each had at least 10 years of experience in evaluating their clients' IAF. The interviews were semi-structured and the list of questions posed is provided in Appendix A. The interviews were conducted by telephone and recorded with the interviewees' permission. The interviews were transcribed and the transcription was validated against the original audio recordings. Interviews were designed to take approximately 30 min and ranged between 17 and 40 min of actual interview time.

While all respondents differed in opinion on the relative importance of each of the factors to their strength judgments, none of them stated that they would follow a particular order in terms of evaluating the evidence. Representative quotes from each of the partners interviewed include the following:

[9] We recognize that our data represent "elicitations" of belief masses rather than theoretical beliefs as defined in Shafer (1976). Even so, we use the shorthand label "beliefs" to simplify the language in this section of the paper.

[10] As in Desai et al. (2010), we consider how the external auditor makes a judgment about the strength of the IAF. In this study, we examine a low materiality/low account level risk area in order to highlight the impact of changes in assessment of W, C and O on the external auditor's strength judgment.

[11] The PCAOB also indicates the reliance decision should be based on inherent risk. The case-based scenario presented to participants in our experiment was designed such that inherent risk was low and thus, standards suggest the external auditor can rely on the work of the IAF if their evaluation of the IAF is strong.

[12] While Desai et al. (2010) argue that the same evidence about corporate governance will affect the perceived quality of W and O, we follow prior research (DeZoort et al., 2001; Maletta, 1993) that includes evidence of the quality of corporate governance into the objectivity factor only.

"*…I don't think there's an order. I think they're all three looked at. Is there a percentage? I don't know. I guess depending on the individual methodology and approach, one person might say work performance is slightly more important or not. To me, they're sort of… you either have them or you don't, and if any one of them's missing, I have a concern. I think they're all there, and it's one of those you either have it or you don't.*"

(Partner 1)

"*That's an interesting question. I don't know that anybody follows a particular order. I think they might certainly weigh the factors a little bit differently … If they're exceedingly competent and can clearly document thought process and all those sorts of things, and they are not independent, then you're in bad shape. Obviously, you can't rely on it… Generally, reviewing work is principally the indicator of how high or low their competence is.*"

(Partner 2)

"*The firm doesn't have a criteria that one of those three trumps the other, but my personal experience is that if I have an internal auditor or an internal audit department…… Typically, if I've got someone who's very competent, that translates usually into very good work performance… I'd say competency directly more relates to work performance than competence relates to objectivity.*"

(Partner 3)

"*Certainly, at our firm, I've never seen any guidance that says 'you should evaluate this in this particular order' and I guess in my own mind, when I evaluate their work, I really do probably take into consideration all three simultaneously although I will say, I may not weight them all the same…it's kind of complex… I guess in a way, I don't mentally think in the certain order.*"

(Partner 4)

"*I wouldn't say that we have a specific order that we follow, no. There's a lot of subjectivity to evaluation of internal audit…we have to look at the control environment as a whole.*"

(Partner 5)

Thus, we conclude that our choice to begin by asking our experimental participants to evaluate evidence about W, while practically important to our operationalization of the relationships examined by Desai et al. (2010), is also reasonably consistent with (or at least not contrary to) the method adopted for evaluating the strength of the IAF in practice.

Second, Desai et al. (2010) assert that the interrelationships between W, C, and O are bidirectional in nature. In the case of r_1, for example, if there is a belief, say 0.9, that the internal auditor is competent, then work performance will be satisfactory with a belief 0.9^*r_1 in absence of any other evidence. In a similar way, if there is a belief, say 0.8, that the work performance of an internal auditor is satisfactory, then the internal auditor is competent with a belief 0.8^*r_1 again in absence of any other evidence. The other interrelationships operate in a similar fashion. Essentially, the bidirectional nature of the interrelationship implies that the order in which the three factors are evaluated would have no bearing on the strength of the interrelationships.

3.3. Experimental participants

Experimental participants were recruited mainly through personal contacts of the authors. Our screening criteria for participation were U.S.-based auditors with a professional accounting designation, and prior experience evaluating the work of the internal auditor in the context of an external audit of financial information or in evaluating internal control. In total, 135 auditors (93 male, 38 female, four did not identify) volunteered to participate in the study. Our initial data screening revealed eight participants that reported having little to no prior experience evaluating the work of the internal auditor. We exclude these participants from our final sample as they did not meet our screening criteria. Big Four or other large national accounting firms employed the remaining 127 participants. Participant's median age fell within the 31–40 years category and 90% had worked in the field of public accounting for four years or more (39% for 10 years or more). All participants were at the senior accountant level or above, distributed as follows: partners (25.2%), senior managers (31.9%), managers (20.7%), and seniors (17%).[13]

Participants were randomly assigned to one of the eight experimental conditions. About half of the participants completed the experiment on-line after being provided a link to the web site containing the appropriate experimental condition and a password to enter the website. The other half of participants completed a pen-and-paper version of the experiment while attending a training session conducted by their firm. In each case, the task took approximately 20 min to complete.

[13] The remaining 5.2% did not respond to the question about their position within the firm. We ran all hypothesis tests reported later in the paper including age, gender, position in the firm, experience level and mode of completion of the experiment (pen and paper vs. online) as covariates. None of these variables were statistically significant.

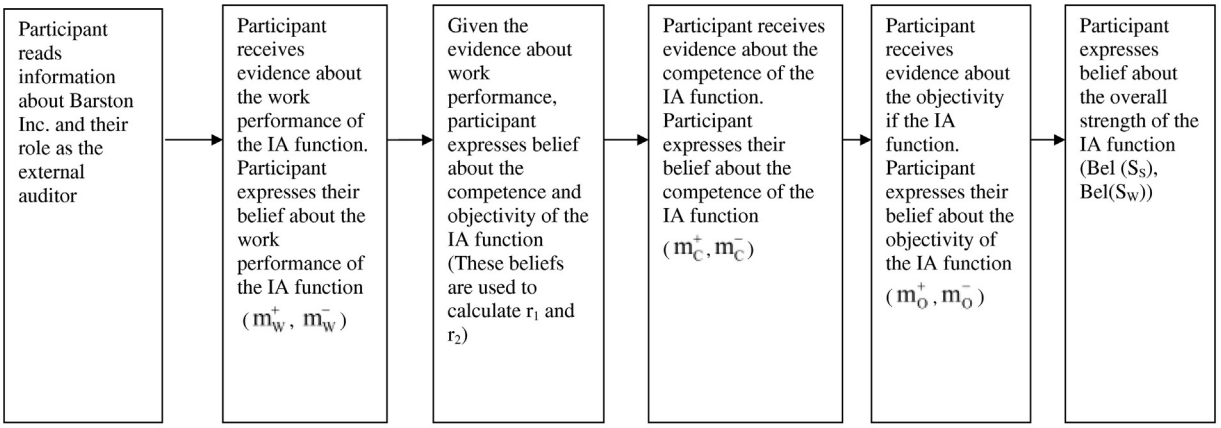


Fig. 2. Timeline of experimental procedures.

3.4. Experimental procedures

Fig. 2 presents a timeline for the judgments made by the participants as they worked through the experimental case. First, the participants read information about Barston Inc. and their role as external auditor. Next, they examine the evidence about internal auditor work performance at Barston Inc. Although the general model developed in Desai et al. (2010) considers cases where a single piece of evidence could provide positive, negative or mixed support, our experimental manipulations contain a set of only purely positive or purely negative evidence about the quality of internal auditor W, C and O. In making this design choice, we were guided by Shafer (1976) who argues that in most real world situations, individual pieces of evidence will tend to be either purely positive or negative in nature. In addition, this design choice simplifies our experimental manipulations and allows us to use Likert scaled items to estimate participant's m-values. Thus, after receiving several pieces of evidence about the IA work performance (W), participants express their belief about the quality of W on a scale of 0 (extremely ineffective) through 10 (extremely effective) with the midpoint of 5 anchored as "neither effective nor ineffective".

Participants then move on to the next page where they are asked to indicate 1) their belief about the competence of the IAF given only what they know about work performance on a scale of 0 (extremely incompetent) through 10 (extremely competent) with the midpoint of 5 anchored as "neither competent nor incompetent" and 2) their belief about the objectivity of the IAF given only what they know about work performance on a scale of 0 (extremely lacking in objectivity) through 10 (extremely objective) with the midpoint of 5 anchored as "neither objective nor lacking in objectivity). This indirect assessment of C and O based only on the evidence about W yields the measure of the strength of interrelationship between W and C (r_1) and W and O (r_2) consistent with the Desai et al. (2010) model. We elicit these beliefs to achieve a key objective of our study; that is, to infer the strength of the judged interrelationships. We then use these beliefs to calculate r_1 and r_2 as in Eqs. (4) and (5) below.[14]

$$r_1 = \frac{\text{Belief in competence given only evidence about work performance}}{\text{Belief in quality of work performance}} \tag{4}$$

$$r_2 = \frac{\text{Belief in objectivity given only evidence about work performance}}{\text{Belief in quality of work performance}} \tag{5}$$

Next, the participants are provided with evidence (either purely negative or purely positive) about the competence (C) of the IAF at Barston Inc. and express their belief about C on a scale of 0 (extremely incompetent) through 10 (extremely competent) with the midpoint of 5 anchored as "neither competent nor incompetent." Participants then receive evidence about the objectivity of the IAF (O) (either purely positive or purely negative) and again express their belief about O on a scale of 0 (extremely lacking in objectivity) through 10 (extremely objective) with the midpoint of 5 anchored as "neither objective nor lacking in objectivity". Finally, participants express their overall belief about the strength of the IAF given all the evidence collected also on a zero through ten scale. Participants then complete manipulation check and demographic questions and log out of the web site or submit the paper copy of the survey.

[14] One advantage of the experimental method is that we can collect evidence in the lab that would be difficult to collect in the field. For example, the auditor's evaluation of the strength of IA work performance evaluated without regard to their competence and objectivity is an important first step in evaluating r_1 and r_2 per the Desai et al. (2010) model but would be unobservable in practice. We therefore end up trading off external for internal validity in that it would be unlikely that auditors in the field would be required to express these beliefs, but we need to collect this data to achieve our research objectives.

4. Results

4.1. Validity and manipulation checks

An initial screening of the data revealed that 18 participants[15] selected a rating of 5 (anchored at "neither effective nor ineffective") in answer to the very first question asked after they read the case scenario: "Based on the evidence provided above, what is your belief about the WORK PERFORMANCE of the internal auditors at Barston Inc.?" Given we had just provided each participant with either all positive or all negative evidence about work performance, this evidence appeared on the same screen (or page) where the rating was elicited and results of extensive pretesting indicated the manipulation of work performance was effective, we suspect these participants were not carefully attending to the information provided in the experimental materials. After excluding these responses, we obtain a final usable sample of 109 participants.[16]

We asked several questions in the post-experimental questionnaire to ensure that the participants understood the experimental manipulations. Specifically, we asked participants their degree of agreement with three statements: "The work performance of the internal auditors at Barston Inc. is satisfactory," "The internal auditors at Barston Inc. are competent" and "The internal auditors at Barston Inc. are objective." Manipulations check questions used a scale of 1 (strongly disagree) to 7 (strongly agree). Mean responses for the participants given the high and low levels of each manipulated variable were significantly different from each other. In addition, mean responses for participants given the high and low levels of each manipulated variable were significantly different from the scale midpoint of 4 ($p < 0.001$) and means for pairs of manipulation check questions are significantly different from each other ($p < 0.001$). Therefore, we evaluate the experimental manipulations as effective.

4.2. Descriptive statistics

To perform our tests, we first convert the scale values provided by the participants for indicating the quality of W, C and O by experimental condition to the form of beliefs that can be used as inputs to the Desai et al. (2010) model (i.e., participants' beliefs about the quality of W (m_W^+, m_W^-), C (m_C^+, m_C^-) and O (m_O^+, m_O^-)). Details about the process of converting scale values to the proper format are provided in Appendix B (Step 1). We report the means (Std. Dev.) of beliefs by experimental condition separated by positive and negative categories in Table 1. These beliefs range in value between zero and one.

Next, we calculate mean r_1 and r_2 in each of the high and low work performance groups using Eqs. (4) and (5) as well as the process described in Appendix B (Step 2). Results are presented in Table 2.[17] A value of zero implies there is no relationship between the two factors and a value of one implies there is a perfect relationship between the two factors. Overall, we find the mean interrelationship between W and C (r_1) to be 0.71 and the mean interrelationship between W and O (r_2) to be 0.55.[18]

4.3. Hypothesis tests

According to H1, external auditors will assign a higher value to r_1 and r_2 when evidence is negative (i.e., when W is judged to be low) than when it is positive (i.e, when W is judged to be high). We test this hypothesis via *t*-tests comparing r_1 when W is high (0.66) versus low (0.76) and r_2 when W is high (0.49) versus low (0.63) (as reported in Table 2). Results indicate that the mean for r_1 is marginally greater when W is low than when it is high ($t = 1.90$, $p = 0.08$, one-tailed) as predicted. These results indicate that, for example, when presented with evidence that W is low, participants form stronger beliefs about the potential for C to also be low than when evidence indicates W is high. This asymmetric relation also holds for r_2. That is, the mean relation r_2 between W and O is marginally higher when evidence indicates W is low ($r_2 = 0.63$) than when W is high ($r_2 = 0.49$, $t = 2.44$, $p = 0.06$, one-tailed). Thus, we find some weak support for H1.

According to H2, when evidence about W, C and O is conflicting, the external auditor's overall belief in the strength of the IAF will be significantly different from the belief derived using the Desai et al. (2010) model. To test this hypothesis, we compare our auditor-participants' overall beliefs in the strength/weakness of the IAF when evidence is conflicting (i.e., some positive and some negative) to those derived from the Desai et al. (2010) model. Model-based beliefs are derived using as inputs the m-values elicited from participants for each of W, C and O as well as our estimates of r_1 and r_2 as inputs to the Desai et al. (2010) model. Results are presented in Table 3.[19]

[15] Of these 18 participants, 8 completed the pen and paper version of the instrument and 10 completed the on-line instrument.

[16] The participants dropped from the sample were distributed relatively evenly across experimental conditions with a minimum of 2 and a maximum of 5 participants dropped from each of the 8 experimental cells. Note we cannot compare results to those obtained when these participants are included in sample because their selection of 5 (neither effective nor ineffective) for W is rescaled in our analysis to a value of 0, and therefore, we cannot calculate the participant's implied r_1 or r_2 since a value of 0 is not logically possible as a denominator in the calculation. Thus, we cannot compare their actual belief in the strength of the IAF with that derived by the model.

[17] We present the mean values for r_1 and r_2 in Table 2 only under the high and low work performance conditions as, consistent with the Desai et al. (2010) model, all participants were required to indicate their beliefs in the internal auditor's competence and objectivity having only evidence that the internal auditors' work performance was either of high or low quality.

[18] According to Desai et al. (2010) and tenets of the belief function formalism, the values of r_1 and r_2 range from zero to one. Our empirical results presented in Table 2 are consistent with this assumption.

[19] Due to our relatively small sample size within some experimental conditions, equivalent non-parametric tests were also performed. Results of the nonparametric tests were qualitatively similar to the parametric tests.

Table 1
Mean (Std. Dev) by experimental condition ($n = 109$).

	Work performance		Competence		Objectivity	
	mw+	*mw−*	mc+	*mc−*	*mo+*	*mo−*
Predominantly negative evidence						
$W_{Low}/C_{Low}/O_{Low}$ ($n = 14$)	0.04 (0.12)	0.40 (0.29)	0.00 (0.00)	0.63 (0.27)	0.00 (0.00)	0.50 (0.26)
$W_{Low}/C_{Low}/O_{High}$ ($n = 12$)	0.08 (0.16)	0.30 (0.26)	0.00 (0.00)	0.58 (0.30)	0.52 (0.30)	0.05 (0.13)
$W_{High}/C_{Low}/O_{Low}$ ($n = 14$)	0.57 (0.17)	0.00 (0.00)	0.00 (0.00)	0.43 (0.23)	0.00 (0.00)	0.63 (0.41)
$W_{Low}/C_{High}/O_{Low}$ ($n = 11$)	0.07 (0.13)	0.31(0.24)	0.56 (0.27)	0.00 (0.00)	0.04 (0.08)	0.69 (0.38)
Predominantly positive evidence						
$W_{High}/C_{High}/O_{High}$ ($n = 16$)	0.62 (0.23)	0.00 (0.00)	0.65 (0.20)	0.00 (0.00)	0.70 (0.21)	0.00 (0.00)
$W_{High}/C_{High}/O_{Low}$ ($n = 18$)	0.66 (0.24)	0.00 (0.00)	0.70 (0.17)	0.00 (0.00)	0.13 (0.24)	0.30 (0.42)
$W_{High}/C_{Low}/O_{High}$ ($n = 15$)	0.51 (0.27)	0.00 (0.00)	0.03 (0.07)	0.35 (0.25)	0.69 (0.21)	0.00 (0.00)
$W_{Low}/C_{High}/O_{High}$ ($n = 9$)	0.04 (0.09)	0.36 (0.30)	0.42 (0.34)	0.04 (0.13)	0.53 (0.33)	0.02 (0.07)

Notes: mw^+ = positive beliefs about IA work performance. mc^+ = positive beliefs about IA competence. mo^+ = positive beliefs about IA objectivity. mw^- = -negative beliefs about IA work performance. mc^- = negative beliefs about IA competence. mo^- = negative beliefs about IA objectivity. All beliefs are stated on a scale ranging from 0 to 1. The belief function framework allows for both positive and negative beliefs where evidence dictates, even within the same scenario.

As can be seen in Table 3 (Panel A), in the three cells where evidence indicates two of the factors are low and one is high, the mean observed belief in the overall weakness of the IAF, Bel (S_w), is significantly different from the mean derived belief in the weakness of the IAF. These results are consistent with our prediction in H2a. On the other hand, the differences in mean and derived beliefs in the strength of the IAF when the evidence is predominantly negative are not significant. We observe that our participants and the model both assign negative beliefs (i.e., the IAF is weak), although the model consistently places more weight on the negative evidence than do the participants.

Next we compare mean observed and derived beliefs in the three cells where evidence indicates two of W, C and O are high and one is low, i.e., results are predominantly positive (Table 3, Panel B). These results are less clear. For example, in the $W_{High}/C_{High}/O_{Low}$ condition, there is a significant difference in observed belief in the weakness of the IAF, Bel (S_w) = 0.10 (Std. Dev. = 0.25) and the mean derived belief in the weakness of the IAF, Bel (S_w) = 0.34 (Std. Dev. = 0.35) ($t = -3.91$, $p < 0.001$) consistent with H2a, but there is no significant difference in the mean observed belief in the strength of the IAF, Bel (S_s) = 0.37 (Std. Dev. = 0.30) and the mean derived Bel(S_s) = 0.31 (Std. Dev. = 0.37) ($t = 0.70$, $p = 0.50$). While W and C are both high in this case, it appears the model weights the negative evidence about objectivity more heavily than do the auditors. We observe that the overall direction of the auditors' beliefs in this cell is positive [i.e., mean observed belief in the strength of the IAF, Bel (S_s) = 0.37, and is significantly higher than mean observed belief in its weakness, Bel (S_w) = 0.10 ($t = 2.36$, $p = 0.03$)] while the overall direction of derived beliefs is indeterminate in this case [i.e., mean derived belief in the strength of the IAF, Bel (S_s) = 0.31 is not significantly different from the mean derived Bel (S_w) = 0.34 ($t = 0.14$, $p = 0.89$)]. Thus, the results derived from the model are consistent with auditing standards in that they suggest caution in the case where O is low, while the auditors would be likely to rate the IAF as relatively strong based on the mean observed beliefs in this cell.

In the two remaining cells presented in Table 3 (Panel B) where O is high, but W and C are either high or low, the derived beliefs in the weakness of the IAS are significantly different from the observed beliefs as predicted in H2. We suspect that the auditors have a more difficult time evaluating combinations of evidence indicating that W is high and C is low or W is low and C is high since, as suggested by Margheim (1986), it would be very rare for auditors to have experienced real world situations where this would be the case. Further it confirms the Libby (1981) finding that accountants have problems integrating evidence to arrive at decisions specifically when evidence is conflicting and diverse in nature.

To complete our analysis, we next examine the conditions in which evidence about W, C and O is either all positive or all negative and thus, W, C and O are consistently high or consistently low. As indicated in Table 3 (Panel A), when W, C and O are all low, the mean observed belief in the strength of the IAF, Bel (S_s), is 0 while the mean derived belief is 0.19 (Std. Dev. = 0.07). These means are not significantly different from one another ($t = -1.33$, $p = 0.21$). On the other hand, the mean observed belief that the IAF is weak in this case, Bel (S_W) = 0.49 (Std. Dev. = 0.27) is significantly lower than the mean derived belief, Bel (S_w) = 0.77 (Std. Dev. = 0.27) ($t = -4.61$, $p < 0.001$). Thus, the model is assigning a higher overall belief in the weakness of the IAF when all of W, C and O are low than are the auditors.

Next, we examine the differences between mean observed and derived beliefs in the cells where the evidence is all positive. As indicated in Table 3 (Panel B), in the cell where evidence indicates W, C and O are all high, there is no statistically significant difference between the mean observed belief in the strength of the IAF Bel (S_s) = 0.66 (Std. Dev. = 0.21) and the mean derived belief Bel (S_s) = 0.73 (Std. Dev. = 0.25) ($t = -1.10$, $p = 0.29$). In addition, the mean observed and derived beliefs in the weakness of the IAF Bel (S_w) in this case are both zero. These results are not surprising given the judgment is much less complex when all evidence is in the same direction and all of it is positive.

4.4. Decision to rely

Although many factors including materiality and risk enter into the reliance decision, we thought it would be interesting to explore how participants' across experimental conditions would respond to a question about potential reliance on the work of

Table 2
Mean (Std. Dev) for values of r_1 and r_2 ($n = 109$).

Evidence about the quality of IA function work performance	Interrelationship between work performance and competence r_1	Interrelationship between work performance and objectivity
High	0.66[a]	0.49[b]
	(0.42)	(0.47)
	$n = 63$	$n = 63$
Low	0.76[a]	0.63[b]
	(0.37)	(0.50)
	$n = 46$	$n = 46$
Total	0.71	0.55
	(0.40)	(0.49)
	$n = 109$	$n = 109$

Each of r_1 and r_2 is calculated using Eqs. (4) and (5) based on beliefs gathered from participants given evidence indicating work performance is high quality or low quality. The theoretical range in value is between 0 and 1 with a higher value indicating a stronger interrelationship. The values r_1 and r_2 are inputs to the Desai et al. (2010) model representing the interrelationships between work performance and competence (R_1) and work performance and objectivity (R_2).

[a] Significantly different from each other at $p = 0.08$ (one-tailed).

[b] Significantly different from each other at $p = 0.06$ (one-tailed).

the IAF at Barston Inc. To do so, we asked our participants to indicate their degree of agreement with the following statement: "Based on ALL the evidence provided to me as the EXTERNAL AUDITOR, I would rely on the work of the internal auditors at Barston Inc." They responded on a scale of 1 (strongly disagree) through 7 (strongly agree). The overall mean response was 3.57 (Std. Dev. = 1.81) and the responses covered the complete theoretical range of 1 through 7. Mean responses were highest in the cell where W, C and O were all high (mean = 5.38, Std. Dev. 0.89) and lowest in the condition where W, C and O are all low (mean = 2.07, std. dev. = 0.99). Overall, the correlation between responses to the reliance question and the strength question (i.e., our dependent variable) was positive and significant ($r = 0.89$, $p < 0.001$) indicating as the judged strength of the IAF increased so did the degree to which the external auditor indicated they would rely on the IA's work.

Out of the responses which were >5 and less than or equal to 7, which represent strong decision to rely, 82% of the responses were in the cells where at least two of the three factors (W, C, and O) were high. Further, 25% of the responses were in the cells where W and C were high and O was low, while around 17% of the responses were in the cells where W and O were high and C was low, and only 11% of the responses were in the cells where O an C were high and W was low. As for the responses which were equal to 4, which represent the participants who were on the fence regarding their decision to rely, 78% of the responses were in the cells where at least two of the factors were high and 22% of the responses were in the cells where at least two of the factors were low.

5. Discussion

In our experiment, participants first read the case scenario and examine the pieces of evidence they can use to evaluate the strength of the company's IAF. Next, they indicate their belief in the quality of work performance, competence and objectivity as well as their belief in the overall strength or weakness of the IAF. In addition, participants form intermediate judgments that allow us to estimate the magnitude of their judged interrelationships between work performance and competence (r_1) and work performance and objectivity (r_2), which prior research argues to be important, but not directly observable.

First, we find the mean interrelationship between work performance and competence (r_1) and between work performance and objectivity (r_2) is stronger when work performance is low than when work performance is high. While these results are inconsistent with the assumption of Desai et al. (2010), they are consistent with our H1 and prior research indicating auditors tend to be more sensitive to negative than positive evidence (Ashton and Ashton, 1988; Butt and Campbell, 1989).

Second, we find when evidence indicates W, C and O are all or predominantly low, observed and derived beliefs are in the same direction (i.e., more negative than positive), but the beliefs derived from the model are consistently more negative than our auditor-participants' beliefs. Contrary to our expectation that auditors will be particularly sensitive to negative evidence (Asare, 1992; McMillan and White, 1993; Knechel and Messier, 1990) as compared to the model, it appears they are not adjusting their beliefs enough in the negative direction based on negative evidence. This is particularly surprising given our finding that negative evidence has a more negative impact on r_1 and r_2 than does positive evidence which is inconsistent with the assumptions of the model. Although our auditor participants are more sensitive to negative than positive evidence when assigning values to r_1 and r_2, they do not appear to be as sensitive to negative evidence when integrating information cues into an overall belief as is the model.

The empirical evidence provided in the current study serves to substantiate the modeled interrelations in prior work. While prior research (Desai et al., 2010; Krishnamoorthy, 2002) has argued in support of considering interrelationships between work performance and competence and between work performance and objectivity, ours is the first study to empirically estimate and examine such relationships in external auditors' assessments of the strength of the IAF. Given the increasing reliance on internal auditors in the financial statement audit and in the evaluation of entity internal controls (SOX 404), a greater understanding of how auditors combine and evaluate these factors into an overall strength judgment is necessary and important.

Table 3
Mean observed and derived overall beliefs in the strength or weakness of the internal audit function by experimental condition ($n = 109$).

	Observed Mean (SD)	Derived Mean (SD)	Test of Differences
Panel A: Predominantly negative evidence			
$W_{Low}/C_{Low}/O_{Low}$ ($n = 14$)			
Bel (S_s)	0.00 (0.00)	0.07 (0.20)	$t = -1.33$ ($p = 0.21$)
Bel (S_w)	**0.49** (**0.27**)	**0.77** (**0.27**)	$t = -4.61$ ($p < 0.001$)
$W_{Low}/C_{Low}/O_{High}$ ($n = 12$)			
Bel (S_s)	0.05 (0.12)	0.19 (0.28)	$t = -2.27$ ($p = 0.06$)
Bel (S_w)	**0.37** (**0.27**)	**0.63** (**0.34**)	$t = -5.44$ ($p < 0.001$)
$W_{High}/C_{Low}/O_{Low}$ ($n = 14$)			
Bel (S_s)	0.01 (0.05)	0.08 (0.21)	$t = -1.19$ ($p = 0.26$)
Bel (S_w)	**0.41** (**0.34**)	**0.73** (**0.32**)	$t = -5.23$ ($p < 0.001$)
$W_{Low}/C_{High}/O_{Low}$ ($n = 11$)			
Bel (S_s)	0.11 (0.19)	0.49 (0.36)	$t = -0.29$ ($p = 0.78$)
Bel (S_w)	**0.13** (**0.28**)	**0.72** (**0.37**)	$t = -2.88$ ($p = 0.02$)
Panel B: Predominantly positive evidence			
$W_{High}/C_{High}/O_{High}$ ($n = 16$)			
Bel (S_s)	0.66 (0.20)	0.73 (0.25)	$t = -1.10$ ($p = 0.29$)
Bel (S_w)	0.00 (0.00)	0.00 (0.00)	No difference
$W_{High}/C_{High}/O_{Low}$ ($n = 18$)			
Bel (S_s)	0.37 (0.30)	0.31(0.37)	$t = 0.70$ ($p = 0.50$)
Bel (S_w)	**0.10** (**0.25**)	**0.34** (**0.35**)	$t = -3.91$ ($p < 0.001$)
$W_{High}/C_{Low}/O_{High}$ ($n = 15$)			
Bel (S_s)	0.25 (0.35)	**0.45** (**0.45**)	$t = -2.25$ ($p = 0.04$)
Bel (S_w)	**0.03** (**0.07**)	**0.19** (**0.22**)	$t = -2.68$ ($p = 0.02$)
$W_{Low}/C_{High}/O_{High}$ ($n = 9$)			
Bel (S_s)	0.29 (0.27)	0.42 (0.35)	$t = -1.20$ ($p = 0.27$)
Bel (S_w)	**0.07** (**0.10**)	**0.25** (**0.26**)	$t = -2.46$ ($p = 0.04$)

Notes: Bel (S_s) is overall belief in the strength of the IAF while Bel (S_w) is the overall belief in its weakness. Observed mean is based on participants' judgments while derived mean is based on the Desai et al. (2010) model. Means in bold indicated significant differences between observed and derived beliefs. W = work performance, C = competence and O = objectivity.

We also note that there is pressure on external auditors from the PCAOB to place an appropriate amount of reliance the work of the IAF where possible (Public Company Accounting Oversight Board, PCAOB, 2009). Our results may suggest that, due to external pressure to find the client IAF to be strong, auditors may be underweighting negative evidence about W, C and O in developing their judgments of IAF strength. This suggests that a decision aid based on the Desai et al. (2010) model could be very helpful in the field.

Due to its intuitive appeal, the diagram provided in Fig. 1 could be used by external audit staff to guide collection of evidence in a structured way, to develop judgments on a scale of 0 to 1 about whether the evidence indicates work performance, competence and objectivity are of high or low quality, and to estimate the strength of the IAF based on these judgments. We view such a decision aid as a potential tool to bring about some consistency and structure to the present relatively unstructured and sometimes inconsistent evaluation of the IAF. We consider the current study to be only the first step in the process of developing such an aid.

Other issues that need to be addressed in future research and that would move us closer to a viable decision aid include the following. First, we developed a relatively simple scenario for experimental purposes where each piece of evidence only provided information about one of the three factors and second, the evidence provided to each participant about each factor was either all positive or all negative. It is likely that in reality the auditor also collects evidence that has both a positive and a negative impact on one of the factors or where one piece of evidence affects judgments of the quality of more than one factor. Thus, future research must consider what would be the effect of increasing the complexity of the decision structure reflected in the Desai et al. (2010) model by loosening these parameters.

Second, we selected a particular order in which to present evidence and elicit participants' beliefs to maintain relative simplicity and clarity in our experimental design. That said whether a different order of presentation would have substantially changed our experimental results is an empirical question that requires follow-up in future research. Third, our results indicate a consistent underweighting of negative evidence by our experimental participants relative to the model. While differences between observed and derived beliefs were predicted when evidence is conflicting, further work to better understand the consistent underweighting of beliefs relative to the model is required.

Finally, other contextual factors that might also affect the final reliance decision include the degree of subjectivity of the task (Glover et al., 2008), the outsourcing or co-sourcing of the IA function (Desai et al., 2011) and the effect of using the internal audit department as a management training ground (Messier et al., 2011). These contextual factors provide an interesting avenue for future work.

Acknowledgements

We would like to thank our auditor participants for their willingness to take part in our experiment and Lan Guo for her help gathering our pre-test data. In addition, we thank Vicky Arnold, Ganesh Krishnamoorthy, Ted Mock, Robin Roberts and workshop participants at Florida Atlantic University, Florida Gulf Coast University and Florida International University for their helpful comments on previous versions of this paper.

Appendix A. Interview questions providing guidance to discussions with audit partners about the process of evaluating the work of their clients'internal audit function

1. Auditing standards require external auditors to evaluate the work performance, competence and objectivity of their client's internal auditors. In theory, in what order should each of these factors be considered when making this judgment?
   In practice, can this order still be followed? What limitations, if any, have you experienced in the practical application of the standard when evaluating the strength of your clients' internal audit function?
2. If the internal auditor's work performance is evaluated as high quality, are you able to make inferences about their competence or objectivity?
   Or would you first have to evaluate competence or objectivity in order to make inferences about work performance?
3. In your opinion, which of the three factors, work performance, competence or objectivity, would you rate as the most important to your evaluation of your clients' internal audit function?
   (Alternatively, maybe it is not possible to select one as more important than the others? If not, could you elaborate?) [Only ask if this seems to be an issue for the respondent]
4. Assuming you found two of the three factors to be strong, but one of them to be relatively weak. Would your strength judgment depend on which one of the three factors was weak? In other words, is any one of work performance, competence or objectivity a particular "deal breaker?"

Appendix B. Sample calculations for observed and derived beliefs

Step 1: Participant reads evidence indicating work performance is high. Participant provides their belief about work performance based on the evidence provided. Assume the participant indicates a belief of 8 out of 10 (i.e., toward the "extremely effective" end of the scale). This is a positive belief as it is greater than the neutral point of 5 on the scale. We rescale this expressed belief as $(8-5)\,/\,5 = 0.6$. We label this belief as $m_W^+ = 0.6$ and infer $m_W^- = 0.0$. We then calculate $m_W^\Theta = (1 - m_W^+ - m_W^-) = 0.4$.

Step 2: Participant next provides their belief in the competence and the objectivity of the IA function based ONLY on the evidence provided so far about work performance. Assume the participant assigns a score of 7 out 10 for competence given evidence about work performance. This score is rescaled as in Step 1 above and indicates a positive belief in IA competence given evidence about work performance of $(7-5)\,/\,5 = 0.4$. Next assume the participant selects a score of 4 out of 10 for objectivity given evidence about work performance. This score is rescaled as in Step 1 above and indicates a negative belief about objectivity given evidence about work performance of $(5-4)\,/\,5 = 0.2$.

Step 3: Next we calculate r_1 and r_2 as follows:

$$r_1 = \frac{\text{Belief in competence given evidence about work performance}}{\text{Belief in quality of work performance}} = 0.4/0.6 = 0.67$$

$$r_2 = \frac{\text{Belief in objectivity given evidence about work performance}}{\text{Belief in quality of work performance}} = 0.20/0.6 = 0.33$$

Step 4: Next we present evidence about competence of the internal audit function and the participant provides a score which is rescaled as in Step 1 above to determine m_C^+ or m_C^- and used to calculate $m_C^\Theta = (1 - m_C^+ - m_C^-)$. Finally, we present evidence about objectivity of the internal audit function and the participant provides a score which is rescaled as in Step 1 above to determine m_C^+ or m_C^- and used to calculate $m_C^\Theta = (1 - m_C^+ - m_C^-)$. We now have all of the inputs to the calculation of $Bel(S_S)$ or $Bel(S_W)$.

Appendix C. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.accinf.2016.12.001.

References

American Institute of Certified Public Accountants (AICPA), 2014,. Using the work of internal auditors. Statement on Auditing Standards No. 128. AICPA, New York, NY.
American Institute of Certified Public Accountants (AICPA), 1991,. The auditor's consideration of the internal audit function in an audit of financial statements. Statement on Auditing Standards No. 65. AICPA, New York, NY.

Abdel-Khalik, R.A., Snowball, D., Wragge, J.H., 1983. The effects of certain internal audit variables on the planning of external audit programs. Account. Rev. 58 (2), 215–227.
Arnold, V., Collier, P.A., Leech, S.A., Sutton, S.G., 2004. Impact of intelligent decision aids on expert and novice decision-makers' judgments. Account. Finance 44 (1), 1–26.
Asare, S.K., Wright, A., 1997. Hypothesis revision strategies in conducting analytical procedures. Acc. Organ. Soc. 22 (8), 737–755.
Asare, S.K., 1992. The auditor's going-concern decision: interaction of task variables and the sequential processing of evidence. Account. Rev. 67 (2), 379–393.
Ashton, A.H., Ashton, R.H., 1988. Sequential belief revision in auditing. Account. Rev. 63 (4), 623–641.
Bame-Aldred, C.W., Brandon, D.M., Messier Jr., W.F., Rittenberg, L.E., Stefaniak, C.M., 2012. A summary of research on external auditor reliance on the internal audit function. Audit. J. Pract. Theory 32 (1), 251–286.
Bell, T.B., Bedard, J., Johnstone, K., Smith, E.F., 2002. KRisk: a computerized decision aid for client acceptance and continuance risk assessments. Audit. J. Pract. Theory 21 (2), 97–113.
Bell, T.B., Carcello, J.V., 2000. A decision aid for assessing the likelihood of fraudulent financial reporting. Audit. J. Pract. Theory 19 (1), 169–184.
Brown, P.R., 1983. Independent auditor judgment in the evaluation of the internal audit functions. J. Account. Res. 21, 444–455.
Butt, J.L., Campbell, T.L., 1989. The effects of information order and hypothesis-testing strategies on auditors' judgments. Acc. Organ. Soc. 14 (5), 471–479.
Clark, M.E., Gibbs, T.E., Schroeder, R.G., 1980. Evaluating internal audit departments under SAS no. 9: criteria for judging competence, objectivity and performance. Woman CPA 22 (July), 8–11.
Curley, S.P., 2007. The application of Dempster-Shafer theory demonstrated with justification provided by legal evidence. Judgm. Decis. Mak. 2 (5), 257–276.
Curley, S.P., Golden, J.I., 1994. Using belief functions to represent degrees of belief. Organ. Behav. Hum. Decis. Process. 58, 271–303.
Dawes, R.M., 1979. The robust beauty of improper linear models in decision making. Am. Psychoanal. 34 (7), 571–582.
DeZoort, F.T., Houston, R.W., Peters, M.F., 2001. The impact of internal auditor compensation and role on external auditors' planning judgments and decisions. Contemp. Account. Res. 18 (2), 257–281.
Desai, N., Gerard, G.J., Tripathy, A., 2011. Internal audit sourcing arrangements and reliance by external auditors. Audit. J. Pract. Theory 39 (1), 149–171.
Desai, V., Roberts, R.W., Srivastava, R., 2010. An analytical model for external auditor evaluation of the internal audit function using belief functions. Contemp. Account. Res. 27 (2), 537–595.
Edge, W.R., Farley, A.A., 1991. External auditor evaluation of the internal audit function. Account. Finance 31 (1), 69–83.
Einhorn, H.J., Hogarth, R.M., 1975. Unit weighting schemes for decision making. Organ. Behav. Human Perform. 13 (2), 171–192.
Fukukawa, H., Mock, T.J., 2011. Audit risk assessment using belief versus probability. Audit. J. Pract. Theory 30 (1), 75–99.
Glover, S.M., Prawitt, D.F., Wood, D.A., 2008. Internal audit sourcing arrangements and the external auditor's reliance decision. Contemp. Account. Res. 25 (1), 193–213.
Gramling, A.A., Maletta, M., Schneider, A., Church, B., 2004. The role of the internal audit function in corporate governance: a synthesis of the extant internal auditing literature and directions for future research. J. Account. Lit. 23, 193–244.
Gramling, A.A., Maletta, M., Myers, P.M., 1997. The perceived benefits of certified internal auditor designation. Manag. Audit. J. 12 (2), 70–79.
Harrison, K., Srivastava, R.P., Plumlee, R.D., 2002. Auditors' evaluations of uncertain audit evidence: belief functions versus probabilities. In: Srivastava, R.P., Mock, T.J. (Eds.), Belief Functions in Business Decisions. Physica-Verlag, New York.
Hurtt, R.K., Brown-Liburd, H., Earley, C.E., Krishnamoorthy, G., 2013. Research on auditor professional skepticism: literature synthesis and opportunities for future research. Audit. J. Pract. Theory 32 (1), 45–97.
International Federation of Accountants (IFAC), 2012. ISA 610 (revised). Using the Work of the Internal Audit Function and Internal Auditors to Provide Direct Assistance.
Knechel, W.R., Messier, W.F., 1990. Sequential auditor decision making: information search and evidence evaluation. Contemp. Account. Res. 6 (2), 386–406.
Krishnamoorthy, G., 2002. A multistage approach to external auditors' evaluation of the internal audit function. Audit. J. Pract. Theory 95–121 (Spring).
Libby, R., 1981. Accounting and Human Information Processing: Theory and Evidence. Contemporary Topics in Accounting SeriesPrentice-Hall.
Maletta, M.J., 1993. An examination of auditors' decisions to use internal auditors as assistants: the effect of inherent risk. Contemp. Account. Res. 508–525 (Spring).
Margheim, L.L., 1986. Further evidence on external auditors' reliance on internal auditors. J. Account. Res. 194–205 (Spring).
McMillan, J.J., White, R.A., 1993. Auditors' belief revisions and evidence search: the effect of hypothesis frame, confirmation bias, and professional skepticism. Account. Rev. 443–465.
Messier, W.F., Reynolds, J.K., Simon, C.A., Wood, E.A., 2011. The effect of using the internal audit function as a management training ground on the external auditor's reliance decision. Account. Rev. 86 (6), 2131–2154.
Messier Jr., W.F., Schneider, A., 1988. A hierarchical approach to the external auditor's evaluation of the internal auditing function. Contemp. Account. Res. 337–353 (Spring).
Payne, J.W., Bettman, J.R., Johnson, E.A., 1993. The Adaptive Decision Maker. Cambridge University Press.
Public Company Accounting Oversight Board (PCAOB), 2009. Report on the first-year implementation of Auditing Standard No. 5. An Audit of Internal Control Over Financial Reporting That is Integrated With an Audit of Financial Statements.
Public Company Accounting Oversight Board (PCAOB), 2007. Auditing Standard No. 5(AS 5) – An Audit of Internal Control Over Financial Reporting That Is Integrated With an Audit of Financial Statements.
Rose, J.M., Rose, A.M., McKay, B., 2007. Measurement of knowledge structures acquired through instruction, experience, and decision aid use. Int. J. Account. Inf. Syst. 8 (2), 117–137.
Rose, J.M., 2002. Behavioral decision aid research: decision aid use and effects. In: Arnold, V., Sutton, S.G. (Eds.), Researching Accounting as an Information Systems Discipline. 5, pp. 111–134.
Russo, J.E., Dosher, B.A., 1983. Strategies for multiattribute binary choice. J. Exp. Psychol. Learn. Mem. Cogn. 9 (4), 676–696.
Schneider, A., 1985. The reliance of external auditors on the internal audit function. J. Account. Res. 911–919 (Autumn).
Schneider, A., 1984. Modelling external auditors' evaluations of internal auditing. J. Account. Res. 657–678 (Autumn).
Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press.
Simon, H.A., 1957. Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in a Social Setting. John Wiley and Sons, Inc., New York.
Srivastava, R.P., Mock, T.J., Turner, T., 2007. Analytical formulas for risk assessment for a class of problems where risk depends on three interrelated variables. Int. J. Approx. Reason. 45, 123–151.
Srivastava, R.P., Mock, T.J., 2002. Belief Functions in Business Decisions. Physical – Verlag, Heidelberg, Springler-Verlag Company.
Srivastava, R.P., Mock, T.J., 2000. Belief function in accounting behavioral research. Advances in Accounting Behavioral Research. 3, pp. 225–242.
Srivastava, R.P., 1993. Belief functions and audit decisions. Audit. Rep. 17 (1), 8–12.
Weisner, M., Sutton, S., 2015. When the World isn't always Flat: The Impact of Psychological Distance on auditors' Reliance on Specialists.