The Impact of Benchmark Set Composition on Auditors' Level 3 Fair Value Judgments

Sudip Bhattacharjee
Virginia Polytechnic Institute and State University

Kimberly K. Moreno
Northeastern University

Nicole S. Wright
James Madison University

ABSTRACT: Auditors frequently use benchmarking analysis to evaluate the appropriateness of a client's estimates. Client management may strategically select benchmark data, making an auditor's evaluation task more difficult. Psychology research suggests that the composition of the benchmark set can lead to contrast effects, because evaluations of an option in a choice set can change based on the inclusion of other options in the set. In Experiment 1, we examine and find that auditors' judgments of the reasonableness of a client-preferred discount rate for a Level 3 investment are inappropriately influenced by the set of peer companies provided by the client as justification. In Experiment 2, audit managers performing the same task similarly exhibited contrast effects. However, as managers' investment experience increased, the influence of contrast effects from the benchmark set decreased. Given the widespread use of benchmark data, contrast effects from benchmark set composition have implications for accounting and auditing contexts.

Keywords: fair value; Level 3 investment; benchmarks; contrast effects.

I. INTRODUCTION

Benchmarking analysis is useful for evaluating the appropriateness of a reporting company's judgments and decisions. However, company managers can strategically select benchmark data, making the evaluation task more difficult. For example, public companies "cherry-pick" peers to justify inflated executive pay to shareholders, masking overpayment to executives (Melin and Diamond 2015). Companies pick the most favorable ratings from sustainability rating providers to make their social responsibility performance appear favorable to investors (Ceres 2011). In addition, managers use strategically selected benchmarks to attain resources for projects that they are championing in internal capital budgeting decisions (Shapiro 2010; Zimmerman 2017).

Of relevance to the current study, auditors frequently use benchmarking analysis to evaluate the appropriateness of a company's estimates or assumptions underlying those estimates. The Public Company Accounting Oversight Board (PCAOB) and the Securities and Exchange Commission (SEC) have recently noted concerns with auditors' assessments of benchmark data when they evaluate inputs of significant assumptions for estimates (Acuitas, Inc. 2017). PCAOB inspection reports have specifically identified deficiencies in auditors' evaluations of the comparability of peer companies that management has used to support their estimates. For example, the PCAOB (2016a, 5) noted in one inspection report that auditors failed to perform procedures to support the reasonableness of management's use of a peer company when developing an independent fair value estimate of restricted shares of common stock. Another report noted that auditors failed to evaluate whether management's use of a peer company's loan portfolio used to justify the classification of debt securities as Level 2 or Level 3 was comparable to

the issuer (PCAOB 2014, 15). The PCAOB (2016b, 14) noted in another report that auditors failed to evaluate whether management's selected attrition rates of peer companies in the industry were comparable to the issuer when auditing the fair value of acquired customer relationship intangible assets. The deficiencies noted in these inspection reports indicate that auditors have difficulty with benchmarking analysis. Given the prevalence of these deficiencies, we examine whether one potential source of these deficiencies is the composition of the benchmark set provided by the client to support the assumptions underlying the estimates. We conducted two experiments to examine whether auditors' evaluations of a client's discount rate for a Level 3 investment are inappropriately influenced by the composition of the client-provided benchmark set to support the discount rate.

Most relevant to the current study, psychology research shows that a decision maker's choice of an alternative in a set changes relative to the composition of the other alternatives in the set (e.g., Bhatia 2013; Huber, Payne, and Puto 1982; Moran and Meyer 2006; Noguchi and Stewart 2014). Contrast effects occur when an alternative is rated more positively when compared to a less positive alternative than when compared to a more positive alternative. This research finds that preferences depend on the comparisons that individuals make among the attributes of the alternatives in the choice set, in addition to the attributes of the alternatives themselves. We propose that if auditors use such comparative processes during benchmarking analysis, then the composition of the benchmark set may inappropriately influence their evaluations. We focus on benchmarking analysis for a Level 3 investment. An auditor uses a variety of inputs, including general economic conditions and the client's own risk factors, when assessing the reasonableness of the discount rate for a Level 3 investment. In addition, one procedure that auditors typically use to help assess the reasonableness of the discount rate is benchmarking analysis. Client management may, however, strategically select the benchmark data they provide to auditors in support of its chosen discount rate for the Level 3 investment.

Consider a fair value scenario in which management compares the discount rate for an investment to two peer companies to support their preferred discount rate. One peer company may have a more conservative rate than the client-preferred rate and have superior similarity to the Level 3 investment, as it is in the same broad industry and has similar product lines. Another peer company may have a more aggressive rate than the rate of the first peer company, but may be moderately similar to the Level 3 investment, as it is in the same broad industry, but the product lines are in a different subsector. In this situation, auditors are likely to identify that the moderately similar peer company is an inferior comparison to the investment relative to the peer company with superior similarity, and will likely recommend a conservative rate. However, a client could add a third peer company that has low similarity to the investment, as its industry is only tangentially related to the investment. Contrast effects research would predict that because the attributes of the third peer company have low similarity to the investment, the moderately similar peer with an aggressive discount rate might appear more similar to the investment. Rather than independently comparing each peer company to the investment, research on contrast effects suggests that auditors may also contrast the low similarity peer company to the moderately similar peer.[1]

Thus, the objective of Experiment 1 is to determine whether contrast effects from the composition of the benchmark set can change the evaluation of the discount rate made by audit seniors when conducting benchmarking analysis. Further, we examine whether very task-specific structured guidance may be successful at reducing contrast effects.

To test this issue, in Experiment 1, audit seniors took the role of an in-charge auditor and assessed a client's discount rate of a Level 3 investment using benchmarking analysis. Management justified the discount rate by providing information on peer companies. Three experimental groups were used for our benchmark set composition manipulation: Two Peer Companies condition, Plausible Third Peer Company condition, and Implausible Third Peer Company condition. In all three conditions, the client identified two peer companies to justify the weighted average cost of capital (WACC) rate: one had superior similarity to the investment with a conservative discount rate, and one had moderate similarity to the investment with an aggressive discount rate. Auditors in the Two Peers condition only received these two peer companies. Auditors in the Plausible Third Peer condition received a third peer company that had low similarity to the investment, which we predict will make the moderately similar peer company appear more similar to the investment due to contrast effects. To rule out that any results derived from the Plausible Third Peer condition were simply due to the presence of a third peer, we created a Third Implausible Peer condition. This group received a third peer company that had such low similarity to the investment that we did not expect auditors to use it in their judgment.

We found that a greater percentage of auditors in the Plausible Third Peer Company condition assessed the discount rate to be consistent with the client-preferred discount rate than auditors in the Two Peer Companies condition, supporting contrast effects. Results from the Implausible Third Peer Company condition suggest that the results are not due to the mere presence of

---

[1] As noted earlier, this benchmarking analysis procedure is only one input in an auditor's judgment of the reasonableness of the discount rate. The auditor will also consider other inputs (e.g., economic and client risk factors) to assess the client's discount rate. However, we focus on benchmarking analysis in this study and hold all other inputs constant across experimental conditions.

a third peer company. We also found that the benchmark set composition had an impact on auditors' documentation of their discount rate assessments, which has far-reaching implications for audit quality. We found that very task-specific instructions reduced, but did not eliminate, the impact of contrast effects due to benchmark set composition.

We ran a second experiment to test whether the results from the first experiment are robust to higher auditor experience levels. Auditors develop knowledge structures through experience (Bonner 2008). These knowledge structures enable auditors to focus on relevant information in a judgment (Shelton 1999; Vera-Muñoz, Kinney, and Bonner 2001; Low 2004). Thus, it is plausible that audit managers may have the knowledge structures necessary to determine which peer companies are most relevant to the benchmarking analysis task, which could reduce the impact of contrast effects. Audit managers in Experiment 2 were randomly assigned to the Two Peer Companies condition or to the Plausible Third Peer Company condition. The experimental materials were identical to those materials used in Experiment 1. Findings from Experiment 2 suggest that managers were susceptible to contrast effects, although they appeared to be less susceptible than seniors. We also examined how managers' task-specific experience (i.e., fair value familiarity and experience in auditing investments) impacts their benchmarking analysis performance. As audit managers' experience in auditing investments increased, the impact of a third plausible peer decreased. The impact of fair value familiarity on contrast effects was not significant. These findings appear to indicate that managers improve their performance in benchmarking as they gain certain task-specific experience.

Client management may strategically choose peer companies with more aggressive assumptions to justify their estimates or the assumptions underlying those estimates. When they do this, our results show that auditors may be inappropriately influenced by the composition of the benchmark set when performing a benchmarking analysis task. Specifically, results from Experiment 1 demonstrate that audit seniors were susceptible to contrast effects from the composition of the benchmark set. Results from Experiment 2 suggest that audit managers were susceptible to contrast effects. However, as audit managers' task-specific experience increased, the influence of the third plausible peer company decreased. Our findings identify potential causes of deficiencies noted by the PCAOB (2014, 2016a, 2016b) inspection reports when auditors assess benchmark data. Our findings should be of interest to audit firms and standard setters when considering the type of guidance and level of auditor expertise needed to improve financial reporting quality of accounting estimates. Our findings are also relevant to any audit area that involves evaluating benchmark appropriateness, such as allowance for loan losses, contingent liabilities, deferred tax assets, goodwill, and hard-to-value investments. These results also have broad implications for other areas of accounting that require benchmarking analysis, such as justification for executive pay, evaluation of social responsibility performance, and support for internal capital budgeting requests.

II. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Deficiencies in Auditing Fair Value

PCAOB inspection reports for large public accounting firms indicate that fair value measurements are the most frequently cited accounts for auditor error (Acuitas, Inc. 2017; Church and Shefchik 2012; Griffith, Hammersley, Kadous, and Young 2015b; International Forum of Independent Audit Regulators [IFIAR] 2016, 2017). Auditors' deficiencies are due, in large part, to the high levels of uncertainty about valuations given volatile financial markets and complex financial instruments (Cannon and Bedard 2017). Importantly, the uncertainty in these estimates can lead to potential management bias in the estimates that poses a risk to financial statement quality. Audit survey research suggests, and PCAOB inspection reports corroborate, that auditors rely too heavily on managements' assumptions in their fair value models (Griffith et al. 2015b; IFIAR 2016). Auditors sometimes fail to adequately test the assumptions underlying managements' models by focusing on only corroborating management assumptions rather than seeking out and incorporating inconsistent data. This potential lack of professional skepticism toward the assumptions of managements' models has captured regulators' attention (Griffith, Hammersley, and Kadous 2015a).

One assumption that auditors may have to evaluate in Level 3 valuations is a client's chosen discount rate. A client can support a particular discount rate for a Level 3 valuation by comparing its discount rate to those of peer companies in the market place (Christensen, Glover, and Wood 2012). However, the PCAOB and SEC have noted recent concerns with auditors' assessments of benchmark data when evaluating the inputs of significant assumptions for estimates (Acuitas, Inc. 2017). As previously indicated, inspection reports specifically note problems with auditors' assessments of the comparability of peer companies to support the reasonableness of significant assumptions (Acuitas, Inc. 2017; PCAOB 2014, 2016a, 2016b).

Prior audit research has examined auditors' assessments of client-provided discount rates for asset valuation in other settings (Earley 2001, 2002). Earley (2001) examined and found that novice auditors could be trained to use information cues of an asset's risk profile to form a judgment about a client's discount rate in a real estate valuation setting. Using the same task, Earley (2002) extended this study by examining whether experienced and novice auditors differ in the way they combine the information cues of a real estate asset to assess the reasonableness of the client's discount rate. The author's results indicate that

experienced auditors outperformed novice auditors in combining multiple information cues to form a judgment for this task. Our study extends this research by examining auditors' judgments in a benchmarking analysis task to assess the reasonableness of a client' discount rate in a Level 3 valuation. We examine whether the composition of the set of peer companies provided by a client to support their discount rate can inappropriately influence auditors' assessments of the rate. We draw on psychology research on comparative decision making to investigate this proposition.[2]

Composition of the Benchmark Set

Psychology research suggests that the choice among a set of options can change with the inclusion of other options in the set (Trueblood, Brown, Heathcote, and Busemeyer 2013). Choice among a set of alternatives has been shown to depend both upon the values of the attributes of the alternatives themselves and upon the comparisons made between the alternatives (Huber et al. 1982; Trueblood and Pettibone 2017). The presence of an alternative provides a reference against which attributes of another alternative can be compared or contrasted. Attribute values of an alternative are used as references in the evaluation of another alternative, which can lead to contrast effects. Research has established that individuals tend to focus more on relative, rather than absolute, differences (Wong and Kwong 2005).

Contrast effects resulting from the composition of a benchmark set provided by a client could occur when auditors evaluate the client's discount rate for a Level 3 investment. One procedure that auditors typically use to help assess the reasonableness of the discount rate is benchmarking analysis. A client may strategically select benchmark data to support a chosen discount rate for the Level 3 investment. The client could compare their discount rate for the investment to a set of peer companies to justify their choice. The auditor may consider several attributes of the peer companies, such as industry, age, and number and type of product lines, to assess whether the peer companies are reasonable benchmarks for the Level 3 investment.

We propose that auditors assess the similarity between a Level 3 investment and a peer company in a holistic manner. That is, industry is likely to be one of the critical attributes an auditor considers in determining the level of similarity between a peer company and an investment. If the investment and the peer company are not in similar industries, then the remaining attributes are unlikely to be relevant because the investment and peer company are not likely to be comparable. For example, consider a scenario where a client provides three peer companies to justify a discount rate assumption. Peer Company A could have superior similarity in terms of how its attributes compare to the investment because it is in the same broad industry as the investment and specializes in similar product lines. Assume that Peer Company A has a conservative discount rate relative to the client-preferred rate. The client could also provide Peer Company B, which has a more aggressive discount rate than Peer Company A and is only moderately similar to the investment relative to Peer Company A because it operates in the same broad industry as the investment, but its product lines are in a different subsector. An auditor's assessment of the client's discount rate could be influenced to be more aggressive and consistent with Peer Company B if the client provides Peer Company C that has low similarity to the investment, as its industry is only tangentially related to the investment. Research has shown that the addition of an option into a choice set can impact the evaluation and choice of another option in the set because it is easy for decision makers to compare options (Noguchi and Stewart 2014). Peer Company C, a plausible but low similarity peer, could change the manner in which auditors assess the appropriateness of the discount rate assumption in the client's model and make the moderate, aggressive Peer Company B appear more similar than it should. This could lead to the auditor accepting a more aggressive discount rate.

Prior research has shown that auditors are susceptible to contrast effects. Auditors' current judgments have been inappropriately influenced by judgments that they have made on a prior client or a prior task (O'Reilly, Leitch, and Wedell 2004; Bhattacharjee, Maletta, and Moreno 2007; Asay, Brown, Nelson, and Wilks 2017). While appropriate comparative data were not available in the prior auditing research on contrast effects, appropriate comparative data readily exist in our study's benchmarking analysis task. We focus on whether auditors do the appropriate comparison (peer company versus investment) or are influenced by additional irrelevant comparisons (peer company versus peer company). Therefore, it is an empirical question as to whether auditors will be susceptible to contrast effects from the composition of the benchmark set in this context. Given that auditors have been found to be susceptible to contrast effects, we propose that even when an appropriate benchmark is provided, an inappropriate, but plausible, benchmark may impact auditors' judgments. In Experiment 1, we examine the impact

---

[2] We conducted a short survey to better understand an audit senior's role in auditing discount rates in Level 3 fair value instruments. Thirty-six audit seniors from Big 4 and national accounting firms with an average of 36 months of audit experience participated in the survey. Auditors responded to questions regarding the frequency of their involvement with auditing Level 3 fair value instruments and involvement in the testing of discount rates (1 = never, 9 = always). When asked about their involvement with auditing Level 3 fair value instruments, 38.9 percent of auditors' responses were in the lower tercile of the scale (1 to 3), 36.1 percent in the middle tercile (4 to 6), and 25 percent in the upper tercile (7 to 9). When asked if they were involved in the testing of discount rates, 38.9 percent of responses were in the lower tercile, 19.4 percent in the middle tercile, and 41.7 percent in the upper tercile. The survey results suggest that senior-level auditors are involved in testing Level 3 fair value instruments and discount rates specifically.

of contrast effects on senior auditors' evaluations of the client's discount rate for a Level 3 investment. The following hypothesis is proposed:

H1: When presented with a third plausible peer company, audit seniors will assess the discount rate to be more consistent with the client-preferred rate than when the third peer company is not presented.

Exploring the Role of Structured Guidance

Given auditors' documented deficiencies in benchmarking analysis, we explore the role that structured guidance could play in improving audit seniors' assessments of client-provided discount rates. An intervention to overcome contrast effects should address the psychological processes underlying the phenomenon. Contrast effects occur because decision makers compare the available alternatives rather than assessing the value of each alternative based on its attributes alone (Noguchi and Stewart 2014). This effect is largely a perceptual effect (e.g., Dhar and Simonson 2003), whereby an alternative in the benchmark set impacts the evaluation of another alternative. Psychology research suggests that interventions that minimize comparisons among alternatives may attenuate contrast effects (Palmer, Maurer, and Feldman 2002; Carlson and Bond 2006). Thus, structured guidance that explicitly instructs auditors to compare the attributes of the peer companies *to the actual investment* when conducting benchmarking analysis may be effective at reducing contrast effects. The potential to improve auditor judgment leads us to explore the role of structured guidance.

III. EXPERIMENT 1

Methodology

Participants

One hundred eleven auditors from Big 4 and other large public accounting firms participated in Experiment 1.[3,4] The auditors were primarily audit seniors (89.2 percent), with an average of 36.69 months of experience, and 87.9 percent worked for a Big 4 accounting firm. Audit seniors often evaluate assumptions related to complex estimates (Griffith et al. 2015a). The participant experience in this study is similar to prior research that has examined the audit of complex estimates utilizing audit seniors (e.g., Griffith et al. 2015a). Auditors were divided into six experimental conditions. Three experimental groups were used to test H1 based on our benchmark set composition manipulation: Two Peer Companies condition, Plausible Third Peer Company condition, and Implausible Third Peer Company condition. To explore the role of structured guidance, three additional experimental groups based on this peer company manipulation received structured guidance. There were no significant differences in months of audit experience across the experimental conditions ($p = 0.65$).

Experimental Procedures

Auditors took the role of an in-charge auditor for a hypothetical client and assessed the discount rate of an investment using benchmarking analysis. Participants read client background information, including information on the management team and select financial information. This information was followed by a discussion of the fair value issue. Participants received a memo prepared by management indicating an investment in a non-public company (Healthcare Innovations), which was still in its product development stage. They received background information and the previous two years of unaudited financial statements for Healthcare Innovations. Management had elected to account for the investment under the fair value method as defined by Accounting Standards Codification (ASC) 820. Due to the lack of identifiable similar investments, management classified the investment as a Level 3 and utilized the discounted cash flows method to assess the fair value of the investment. Auditors were told that the audit team had completed some audit work on the discounted cash flows, and that their task was to assess the weighted average cost of capital (WACC) assumption used by management.

The memo then stated that management recognized an unrealized gain of $9.6 million for the appreciation of their ownership in the investment in the fourth quarter of the current year. Management provided a discounted cash flow worksheet

---

[3] This research was supported by the Center for Audit Quality (CAQ) Access to Audit Personnel program. In addition, some data were gathered from auditors in public accounting firms using personal contacts at a later date. Therefore, while all participants were subject to randomization, data gathering occurred in two stages. Of the 111 audit seniors who were used in the final analysis for Experiment 1, 82 completed the case electronically, while 29 participants completed the instrument on paper during an internal firm training attended by one of the researchers. In total, 46 auditors were excluded from Experiment 1. Forty-five participants clicked on the Qualtrics link, but did not provide any responses, while one auditor only answered the first question before quitting.

[4] The Institutional Review Board at the affiliated university approved the use of human participants for the experiments reported in this paper.

that had been audited by the audit team that detailed the cash inflows and outflows. Finally, management provided support for the WACC assumption used to discount the cash flow projections. The memo summarized the risk profile of Healthcare Innovations using four key attributes of the company—the industry (biomedical technology industry), age (established in 2011), number of product lines (one product line, needleless injectors), and size ($8.4 million total assets). Management stated that they believed an 11 percent WACC was reasonable and provided support for that rate. This support was provided in the form of an industry WACC range (9–15 percent) and information on attributes and the WACC used by different peer companies. These peer companies were used to manipulate the benchmark set composition (see Table 1 for details on the investment and peer companies).

Manipulation of the Benchmark Set Composition

In all conditions, the client had identified two peer companies to justify the WACC rate. The benchmarking analysis task involved comparing the investment to the peer companies with respect to industry, product lines, age, and asset size. In designing profiles for the peer companies, we first focused on the similarity of the investment to the peer companies with respect to industry. The industry in which the companies operate is likely to be the primary determinant of similarity when auditors examine comparability holistically. The first peer company, Anacor, Inc., was designed to have superior similarity to the investment because it was in the same broad industry (biotechnology industry, specifically, medical devices) as the investment (see Table 1). It had a similar age (established in 2009), had a similar number and type of product lines (three product lines, one of which was similar to the investment), and had a similar asset size ($6 million total assets). It utilized a 14.5 percent WACC, a more conservative discount rate than the client's investment. The second peer company, Bivanir Corp., was designed to have moderate similarity to the investment because it was in the same broad industry (biotechnology industry, specifically, hospital equipment) as the investment, but its products (12 product lines) were in a different subsector. It utilized a WACC rate of 10.5 percent, a more aggressive discount rate than the company's investment. We designed these two peer companies such that, without the presence of any additional peer companies, auditors should be able to identify Anacor, Inc. as having superior similarity to the investment and, as such, be skeptical of the client's WACC. Auditors in the Two Peer Companies condition received only these two peer companies.

For the Plausible Third Peer Company condition in Table 1, auditors received information from the client on a third peer company, MedServices Corporation, which was lower in similarity to the investment than Bivanir Corp., the moderate similarity peer. The third peer company was in a tangentially related industry (healthcare and hospital industry). The peer had a WACC rate of 9 percent, which was close, albeit more aggressive than the moderate peer company. Importantly, the purpose of the third plausible peer company was to make the moderate peer company appear more similar to the investment because the third plausible peer, MedServices Corp., had lower similarity to the client's investment than Bivanir Corp. We created an Implausible Third Peer Company condition to rule out the possibility that our results from the Plausible Third Peer Company condition were simply due to the presence of a third peer company. The third implausible peer company, OAI Industries, had a 9 percent WACC rate and had such low similarity to the investment that we did not expect auditors to use it in their judgment.[5]

We gathered data to understand if the third implausible peer company (OAI Industries) was perceived as having lower similarity with the client's investment than the third plausible peer (MedServices Corp.). Ninety-six graduate accounting students with an average of 7.53 months auditing experience (an average of 1.46 auditing internships and 1.08 busy seasons) were randomly assigned to one of two groups. In one group ($n = 48$), the participants were presented with information about the client's investment and the third plausible peer company (MedServices Corp.). They were then asked to rate how comparable the peer company was with respect to its industry, age, number of product lines, and size with the client's investment. They also provided an overall assessment of the comparability of the peer company with the client's investment. In the second group ($n = 48$), participants were presented information on the client's investment and the third implausible peer company (OAI Industries) and were asked to provide the same ratings. Participants rated greater overall similarity (1 = not comparable at all; 9 = very comparable) between the third plausible peer company and the client's investment (mean = 3.15) than between the third implausible peer and the client's investment (mean = 2.48; $t_{94} = 7.715$; $p = 0.003$). Similar results were observed when comparing the client's investment and the third plausible peer versus the client's investment and the third implausible peer with respect to the industry (mean = 4.31 versus 3.13; $t_{94} = 15.78$; $p = 0.001$), age (mean = 3.15 versus 2.00; $t_{94} = 18.74$; $p = 0.001$), number of product lines (mean = 2.42 versus 2.06; $t_{94} = 1.83$; $p = 0.0895$), and size of the companies (mean = 3.15 versus 2.48; $t_{94} = 4.64$; $p = 0.017$). Collectively, these results confirm that participants perceived that the third implausible peer company had lower similarity to the client's investment than the third plausible peer.

---

[5] We gathered open-ended data from an expert panel of four auditors from Big 4 audit firms (three audit partners and one senior manager) to confirm the relevance of this issue in practice. Overwhelmingly, their responses confirm the premise that (1) the use of benchmark comparisons is prevalent in auditing, (2) clients sometimes provide less comparable peer companies, and (3) auditors have deficiencies in assessing peer companies.

TABLE 1

Investment and Peer Company Details and Overview of Benchmark Set Composition of Experimental Conditions

Company Attribute Factors	Investment: Healthcare Innovations	Peer Companies			
		Superior Similarity: Anacor, Inc.	Moderate Similarity: Bivanir Corp.	Plausible Low Similarity: MedServices Corp.	Implausible Low Similarity: OAI Industries
Industry	Biomedical technology industry	Biotechnology industry; specifically, medical devices	Biotechnology industry; specifically, hospital equipment	Healthcare and hospital industry	Consumer Health
Age	Established 2011	Established 2009; Public 2012	Established 2008; Public 2011	Established 1993; Public 1999	Established 1936; Public 1955
Product Lines	One product line—needleless injectors	Three product lines, one of which is similar to Healthcare Innovations' injectors	12 product lines	15 product lines that include pharmaceuticals and medical supplies	Over 90 products within five segments that include Healthcare, Home Care, Family Care
Total Assets	$8.4 million	$6 million	$15 million	$45 million	$1.6 billion
WACC	11% client-provided	14.5%	10.5%	9%	9%

All participants were provided the following information: Investment (Healthcare Innovations) Information, and Bloomberg Biotechnology Industry 2013 WACC range = 9–15 percent.
Experimental Conditions: Benchmark Set Composition:
Two Peer Companies Condition = Superior Similarity (Anacor) and Moderate Similarity (Bivanir) peer companies;
Plausible Third Peer Condition = Superior Similarity (Anacor), Moderate Similarity (Bivanir), and Low Similarity/Plausible (MedServices) peer companies; and
Implausible Third Peer Condition = Superior Similarity (Anacor), Moderate Similarity (Bivanir), and Low Similarity/Implausible (OAI) peer companies.

American Accounting Association

To ensure the veracity of the benchmark data, participants in all conditions were told that a search had been performed by an audit team member and that no additional peer companies were identified, and that the WACC rates and features provided for the peer companies agreed with publicly available information.

Manipulation of Structured Guidance

To explore the role of structured guidance on audit seniors' judgments, the guidance manipulation was first introduced when auditors were told that the task was to assess the discount rate. It again occurred when information was provided on the peer companies, but prior to auditors making the fair value assessment. The structured guidance was designed from an understanding of the mechanisms of contrast effects from psychology research (e.g., Palmer et al. 2002; Carlson and Bond 2006; Noguchi and Stewart 2014). Psychology research suggests that the key to overcoming contrast effects is to get individuals to assess attribute values based on the values themselves and not based on a comparative process. For the structured guidance, participants were told to assess the WACC assumption and related inputs of the client to ensure that they were reasonable, not inconsistent with market information, and based on relevant information. Auditors were then provided very task-specific instructions so that they would compare the peer companies to the investment. Auditors were then given explicit step-by-step instructions to focus on the attributes of the investment relative to those of the peer companies. Participants in the three main experimental conditions designed to test H1 did not receive this guidance. Rather, they were simply told to assess the WACC assumption and related inputs to ensure compliance with ASC 820. All participants documented their assessment of management's WACC assumption.

Dependent Variables and Post Experimental Questions

Thereafter, auditors were asked to list the WACC rate that they believed should be used and provide a justification for that rate. A table was provided that allowed auditors to calculate the unrealized gain or loss and the fair value based on the WACC selected. The relationship between auditors' WACC assessments and unrealized gains or losses and resulting fair value assessments were mechanically derived from the WACC. Auditors then responded to several post-experimental, manipulation check, and demographic questions.[6]

Results

Test of H1—Auditors' Discount Rate Assessments

Table 2 provides results for the test of H1. Panel A of Table 2 reports the frequency (percentage) of auditors' assessed discount rates across the experimental conditions. The client stated that they believed an 11 percent WACC was reasonable. Auditors' responses were classified into Client-Preferred WACC (indicating less than or equal to 11 percent) or More Conservative WACC (indicating greater than 11 percent). Given the structure of our case, a minor change in the discount rate can have a significant impact on the fair value assessment. If auditors disagreed with management's rate of 11 percent, and thought anything greater than 11 percent would be a more reasonable rate, then an adjustment would be recommended to the client in order to issue a clean opinion.[7]

Panel B of Table 2 indicates that 64.7 percent of the auditors in the Plausible Third Peer condition assessed that the discount rate should be less than or equal to 11 percent (similar to what that the client believed was reasonable), as compared to only 5.9 percent of the auditors in the Two Peer Companies condition (Fisher's Exact Test p-value $= 0.001$, one tailed).[8] In our experiment, this difference in the assessed discount rate translates into a large and material difference in the mean unrealized gain assessment between the Two Peer Companies condition ($3,273,213) and the Plausible Third Peer condition ($8,259,488).

We performed an additional analysis to see how many auditors agreed with the superior peer company. Recall that the superior peer company utilized a 14.5 percent WACC. Thus, a greater than or equal to 14 percent category indicates an agreement with the superior benchmark.[9] Interestingly, none of the auditors in the Plausible Third Peer condition assessed that

---

[6] The final versions of the experimental materials were reviewed by eight partners from the audit practice of the firms that participated in the CAQ program. This final review ensured that terminology was consistent with practice and that the experimental task was appropriate for audit seniors. In addition, the partners agreed that the Level 3 fair value task and information provided by management in support of their assumption of the discount rate was realistic.

[7] Recall that auditors were divided into six experimental conditions. Three experimental groups were used to test H1 (Table 2), while three additional experimental groups received structured guidance (Table 4). Across these six experimental conditions, untabulated results also showed that the experimental manipulations significantly influenced the frequency of auditors' assessed discount rates classified into the two WACC categories ($\chi^2 = 31.03$; df $= 5$; p $= 0.001$).

[8] Two-tailed p-values are reported unless otherwise stated. One-tailed values are reported for directional predictions.

[9] Across the three experimental conditions for H1, the vast majority of auditors (92.1 percent) chose whole numbers when indicating their WACC assessment. As such, 14 percent was chosen as the cutoff point for this category.

TABLE 2

EXPERIMENT 1

Impact of Benchmark Set Composition on Audit Seniors' Discount Rate Assessments

Panel A: Frequency (Percentage) of Audit Seniors' Assessments of the WACC by Group[a]

Variable	Two Peer Companies (n = 17)	Plausible Third Peer (n = 17)	Implausible Third Peer (n = 21)
Client-Preferred WACC: Less than or Equal to 11%	1 (5.9%)	11 (64.7%)	2 (9.5%)
More Conservative WACC: Greater than 11%	16 (94.1%)	6 (35.3%)	19 (90.5%)
Total	17 (100%)	17 (100%)	21 (100%)

Panel B: Test of H1[b]

Comparisons	Fisher's Exact Test p-value
Two Peer Companies versus Plausible Third Peer	0.001

Panel C: Test of the Effects of an Implausible Third Peer Company

Comparisons	Fisher's Exact Test p-value
Two Peer Companies versus Implausible Third Peer	0.999

[a] Panel A reports the number of audit seniors, across conditions, that assessed different WACC rates that should be used for the discounted cash flows. Management stated that they believed an 11 percent WACC was reasonable, while greater than 11 percent WACC was more conservative.
[b] Panels B and C report Fisher's Exact Test p-values. The Panel B p-value is one-tailed due to directional predictions, while the Panel C p-value is two-tailed due to non-directional predictions.

the discount rate should be closer to the superior benchmark, while 82.4 percent of the auditors in the Two Peer Companies condition made that assessment ($p < 0.001$). These findings provide support for H1. When the client provided information about a plausible third peer company that made the moderately similar company appear more similar to the investment, more auditors assessed a lower discount rate than when the plausible third peer company was not provided.

To rule out that our results for H1 were due to the mere presence of a third peer company, Panel C of Table 2 tests auditors' assessments of the WACC in the Implausible Third Peer condition to the Two Peer Companies condition. There were no significant differences in the percentage of auditors who assessed that the discount rate should be less than or equal to 11 percent between the auditors in the Implausible Third Peer condition (9.5 percent) and auditors in the Two Peer Companies condition (5.9 percent) (Fisher's Exact Test p-value = 0.999). Furthermore, while 64.7 percent of the auditors in the Plausible Third Peer condition assessed that the discount rate was similar to what the client believed was reasonable, only 9.5 percent of auditors in the Implausible Third Peer condition did so (Fisher's Exact Test p-value = 0.001).[10]

Auditors also rated how reasonable it was for management to use each peer company as a comparison company to assess the client's investment (1 = not reasonable at all; 7 = very reasonable). This variable can provide insight into how the third peer company changed the assessment of the moderately similar peer company. The moderate similarity peer company was assessed higher by auditors in the Plausible Third Peer condition (mean = 4.53) than in the Two Peer Companies condition (mean =

[10] Given the complex nature of the Level 3 fair value assessment, when asked to describe the next step they would take regarding the audit work of the investment, the majority of the auditors decided that they would not make a final conclusion without talking to their manager. This is consistent with the findings of Griffith et al. (2015a), and suggests that auditors at this rank would likely involve their manager in this decision. For each of the three experimental conditions for H1, the following percentage of auditors indicated that they would contact their managers: Two Peer Companies (60 percent), Plausible Third Peer (81.3 percent), and Implausible Third Peer (57.1 percent).

2.18) and the Implausible Third Peer condition (mean = 3.33; both p < 0.5). Thus, contrast effects due to the presence of a plausible third peer company made the moderate (aggressive) peer company appear reasonable.[11]

Auditor Documentation

All participants documented their assessments of management's WACC position. We analyzed this documentation as a second measure of auditors' agreement with the appropriateness of the assumptions used in the client's model. Two authors, blind to the experimental conditions, examined each auditor's documentation. Each assessment of management's WACC position was examined in its totality to ascertain if auditors agreed (or disagreed) with the peer companies provided to justify the discount rate, and if the peers convinced the auditor that management's WACC was reasonable.[12] The agreement level between the coders was 89.2 percent (Kappa = 0.825, p < 0.001) and all differences were reconciled.

The documentation results are reported in Table 3. Panel A of Table 3 reports the number of auditors, across conditions, that agreed or disagreed that the peer comparisons were appropriate or made no mention of them. Untabulated analyses showed that the experimental manipulations significantly influenced auditors' documentation decisions (χ^2 = 12.71; df = 4; p = 0.013). Tests in Panel B indicate that more auditors in the Plausible Third Peer condition agreed that the benchmark set was appropriate (52.9 percent) than auditors in the Two Peer Companies condition (17.6 percent) (Fisher's Exact Test p-value = 0.038).[13] The documentation results suggest that contrast effects due to the peer company set composition may be far-reaching, as this documentation could impact superiors' judgments and even specialists ([Ricchiute 1999]). We also observe that auditor documentation is not susceptible to contrast effects when an implausible third peer company is presented. As reported in Panel C, only 14.3 percent of the auditors in the Implausible Third Peer condition agreed with management's rate, and this level was similar to that of the Two Peer Companies condition (17.6 percent, p > 0.10).

Structured Guidance

We explored whether very structured guidance that provides task-specific instructions can improve audit seniors' assessments of a client-provided discount rate.[14] As reported in Table 4, there was still a significant difference in the percentage of auditors who assessed that the discount rate should be less than or equal to 11 percent between the Plausible Third Peer—Structured Guidance condition (28.6 percent) and the Two Peer Companies—Structured Guidance condition (0.0 percent) (Fisher's Exact Test p-value = 0.027). No significant differences were observed in the percentage of auditors who assessed that the discount rate should be less than or equal to 11 percent between the Implausible Third Peer condition—Structured Guidance

---

[11] Auditors' reasonableness ratings of the superior peer company were similar for the Two Peer Companies (mean = 5.35), Plausible Third Peer (mean = 5.94), and Implausible Third Peer (mean = 6.05) conditions (p > 0.5). The assessment of the third peer company was similar between the Plausible Third Peer (mean = 2.76) and the Implausible Third Peer (mean = 2.33) conditions (p > 0.5). Since the ratings of the plausible and implausible low similarity companies did not differ, we examined how the plausible low similarity company was viewed subject to auditors' discount rate assessments. In the Plausible Third Peer condition, auditors who assessed that the discount rate should be less than or equal to 11 percent rated the plausible third peer company to be more reasonable (mean = 3.45) than auditors who assessed the discount rate to be greater than 11 percent (mean = 1.50, p < 0.05). Auditors who were impacted by the plausible third peer rated this peer company higher than auditors who were not influenced by the company.

[12] As an example, the following documentation by auditors was classified as agreeing with the peer companies provided by the client (excerpts verbatim from auditors' responses): "The WACC falls within the range of comparable companies in an appropriate industry—Given the assets held and the composition of the Company, the level of risk is in line with the WACC of its competitors (i.e. it falls between the asset composition of Bivanir and Anacor, as does the WACC)." "Reasonable: The Company's WACC appears to be a mix between Anacor and Bivanir's WACC's, which seems reasonable due to the similar industry, asset size, product lines, and date of establishment. Consistent: Based on the audit team's search of additional comparative information and verification of comparative WACC's used in the range, I am comfortable that the WACC range should fall within 9 percent–15 percent. Relevant: The risk profile that was used in calculating the Company's WACC seems to be relevant." The following documentation was classified as disagreeing with the peer companies provided by the client (excerpts verbatim from auditors' responses): "As opposed to the comparative companies, which each have multiple product lines in production, Healthcare Innovations, Inc. is 100 percent reliant on the success of the MediJet product line and has currently only obtained one licensing agreement. Considering the concentration of risk on the one product line as compared to Anacor and Bivanir, a WACC heavily weighted on the low-end of the 9–15 percent Bloomberg range seems aggressive. A WACC more in line with Anacor's 14.5 percent seems more appropriate." "Bivanir and MedServices appear to be more established than Healthcare Innovations, with more assets, and more diversified business lines. Also they are not in same industry. I would challenge just how comparable these companies are to Healthcare innovations."

[13] Given that some cells in Panel A of Table 3 have a frequency of less than 5, Fisher's Exact Test was utilized to test the comparisons between groups. Separate 2 × 2 Fisher's Exact Tests were computed for the analyses reported in Panels B and C of Table 3, excluding audit seniors who provided no mention of the peer companies.

[14] All participants were asked to rate the specificity of the guidance they received (1 = not specific at all; 7 = very specific) and how helpful the guidance was (1 = strongly disagree; 7 = strongly agree) when assessing management's WACC position. Auditors in the structured guidance conditions assessed the specificity of the guidance higher (mean = 6.30) and the guidance as more helpful (mean = 6.23) than the specificity (mean = 1.95) and helpfulness (mean = 2.28) assessments of auditors who did not receive the structured guidance (groups used to test H1) (t_{109} = 34.35, p = 0.001; and t_{108} = 27.18, p = 0.001, respectively).

TABLE 3

EXPERIMENT 1

Audit Seniors' Documentation of Agreement with the Peer Companies

Panel A: Frequency (Percentage) of Audit Seniors Agreeing or Disagreeing that the Peer Company Comparisons were Appropriate[a]

Variable	Two Peer Companies (n = 17)	Plausible Third Peer (n = 17)	Implausible Third Peer (n = 21)
Agreed with peer companies	3 (17.6%)	9 (52.9%)	3 14.3%
Disagreed with peer companies	10 (58.9%)	5 (29.5%)	17 81.0%
No mention of peer companies	4 (23.5%)	3 (17.6%)	1 4.8%
Total	17 (100%)	17 (100%)	21 (100%)

Panel B: Test of H1[b]

Comparisons	Fisher's Exact Test p-value
Two Peer Companies versus Plausible Third Peer	0.038

Panel C: Test of the Effects of an Implausible Third Peer Company

Comparisons	Fisher's Exact Test p-value
Two Peer Companies versus Implausible Third Peer	0.658

[a] Panel A reports an analysis of the number of audit seniors, across conditions, that agreed or disagreed that the client's peer company comparisons were appropriate.

[b] Given that some cells in Panel A have a frequency of less than 5, Fisher's Exact Test was utilized to test the comparisons. Separate 2×2 Fisher's Exact Tests were computed for the analyses reported in Panels B and C, excluding audit seniors who provided no mention of the peer companies. The Panel B p-value is one-tailed due to directional predictions, while the Panel C p-value is two-tailed due to non-directional predictions.

condition (10.5 percent) and auditors in the Two Peer Companies condition—Structured Guidance condition (0.0 percent) (Fisher's Exact Test p-value = 0.489).

We next examined if the pattern of the results in the structured guidance conditions are different than the results without the structured guidance. A comparison between the plausible third peer conditions with or without structured guidance indicates that 64.7 percent of the auditors in the Plausible Third Peer condition (Table 2) assessed that the discount rate should be less than or equal to 11 percent, as compared to only 28.6 percent of the auditors in Plausible Third Peer—Structured Guidance condition (Table 4) (Fisher's Exact Test p-value = 0.048). Thus, it appears that very structured guidance providing task-specific instructions can help to partly reduce contrast effects.

To provide insight into the processing mechanism driving the partial success of the structured guidance, 32 additional auditors were randomly assigned to one of two conditions—Plausible Third Peer—No Structured Guidance (n = 17) or Plausible Third Peer—Structured Guidance (n = 15). Auditors received the same experimental materials as auditors in the main experiment. In addition, these auditors assessed how comparable each of the four attributes (i.e., industry, age, number of product lines, and size) of the moderate similarity peer company was to the four attributes of the investment (1 = not comparable at all; 7 = very comparable). The manipulation check and the WACC results are consistent with those reported in the main results in Experiment 1. In addition, auditors' mean ratings of the moderate peer company's industry (6.00), age (6.18), product lines (5.12), and size (4.76) in the Plausible Third Peer—No Structured Guidance were more comparable to the investment than the respective age (4.40), industry (4.93), product lines (2.27), and size (1.40) of the Plausible Third Peer—

TABLE 4

EXPERIMENT 1
Impact of Benchmark Set Composition on Audit Seniors' Discount Rate Assessments
Exploring the Role of Structured Guidance

Panel A: Frequency (Percentage) of Audit Seniors' Assessments of the WACC by Group[a]

	Structured Guidance		
Variable	**Two Peer Companies (n = 16)**	**Plausible Third Peer (n = 21)**	**Implausible Third Peer (n = 19)**
Client-Preferred WACC: Less than or Equal to 11%	0 (0.0)%	6 (28.6)%	2 (10.5)%
More Conservative WACC: Greater than 11%	16 (100)%	15 (71.4)%	17 (89.5)%
Total	16 (100%)	21 (100%)	19 (100%)

Panel B: Test of the Effects of a Plausible Third Peer Company[b]

Comparisons	**Fisher's Exact Test p-value**
Structured Guidance—Two Peer Companies versus Plausible Third Peer	0.027

Panel C: Test of the Effects of an Implausible Third Peer Company

Comparisons	**Fisher's Exact Test p-value**
Structured Guidance—Two Peer Companies versus Implausible Third Peer	0.489

[a] Panel A reports the number of audit seniors, across conditions, that assessed different WACC rates that should be used for the discounted cash flows. Management stated that they believed an 11 percent WACC was reasonable, while greater than 11 percent WACC was more conservative.
[b] Panels B and C report Fisher's Exact Test p-values. The p-values are two-tailed due to non-directional predictions.

Structured Guidance condition (all $p < 0.05$). These findings demonstrate that the partial success of the structured guidance is due to the manner in which the guidance alters the evaluation of the peer attributes.

IV. EXPERIMENT 2

Motivation and Methodology

In Experiment 1, senior auditors were susceptible to contrast effects from the client-provided set of peer companies when performing a benchmarking analysis to assess the discount rate. We ran Experiment 2 to test whether the results from our first experiment are robust to higher experience levels. On one hand, audit managers may be less susceptible to contrast effects because auditors acquire highly developed knowledge structures through experience (Shelton 1999; Bonner 2008). Knowledge structures enable decision makers to focus on important problem features and consider relevant information (Hammersley 2006). Experienced auditors with highly developed knowledge structures focus on relevant information and ignore irrelevant information (Shelton 1999; Vera-Muñoz et al. 2001; Low 2004; Hammersley 2006).

For example, Shelton (1999) found that audit managers and partners were less susceptible to the dilution effect than audit seniors in a going concern judgment. The author contends that more experienced auditors have highly developed knowledge structures and use directed strategies to focus on only relevant information. Most relevant to the current study, O'Reilly, Leitch,

and Wedell (2004) found that task experience in loan analysis weakens, but does not eliminate, the impact of contrast effects from a set of prior loans on a current loan assessment under certain conditions. The authors propose that auditors were less susceptible to contrast effects in conditions where auditors' task experience provided a context to evaluate the current loan and created less reliance on the prior comparative data. When auditors had less task experience, they relied more on the prior loans to assess the current loan. Thus, it is plausible that audit managers may have the necessary knowledge structures to determine which peer companies are most relevant in the current study's benchmarking analysis task. Audit managers are likely to have more experience with auditing Level 3 investments and benchmarking analysis than audit seniors. Consequently, managers may have the developed knowledge structures to be able to identify and attend to the most relevant peer companies in the benchmarking analysis task and be less swayed by a plausible third peer company.

On the other hand, the robustness of contrast effects with experienced decision makers has been demonstrated within and outside the auditing context. As previously noted, O'Reilly et al. (2004) found that task experience attenuated but did not eliminate contrast effects in their loan assessment, demonstrating the strength of comparative decision making. In addition, Asay et al. (2017) found that higher rank auditors (i.e., managers and partners) were susceptible to contrast effects due to different financial reporting regimes. Managers and partners' reporting judgments under IFRS were systematically influenced away from the accounting treatment supported by standards from another regime (U.S. generally accepted accounting principles [GAAP]). Auditors who applied current international financial reporting standards (IFRS) guidance judged an impairment reversal as more appropriate under IFRS when they first judged the appropriateness of an impairment reversal under the U.S. standard. Similarly, experienced decision makers in other applied domains have been susceptible to comparative-based decision making. Experienced Olympic judges, attending physicians, and judicial experts have been found to be susceptible to contrast effects (Damisch, Mussweiler, and Plessner 2006; Englich and Mussweiler 2001; Englich, Mussweiler, and Strack 2006; Yeates, O'Neill, Mann, and Eva 2012). In sum, research within and outside of auditing suggests that contrast effects are robust, and experience has only been found to reduce, but not eliminate, the bias in some conditions. This research leads us to test whether contrast effects will impact audit managers' judgments in our benchmarking analysis task.

Thirty-three managers with an average of 77.52 months of audit experience participated in Experiment 2. The managers reported that they spent an average of 58.33 hours over the last year evaluating audit work performed on the valuation of Level 2 or Level 3 assets or liabilities. Auditors were randomly divided into two of the experimental conditions from Experiment 1, the Two Peer Companies condition and the Plausible Third Peer Company condition. Experimental materials for these two conditions were the same as in Experiment 1.

Results

Audit Managers' Discount Rate Assessments

Panel A of Table 5 presents the frequency (percentage) of audit managers' assessed discount rates for the two experimental conditions. Auditors' responses were classified into two categories: Client-Preferred WACC (indicating less than or equal to 11 percent) or More Conservative WACC (indicating greater than 11 percent). Panel B of Table 5 indicates that 25 percent of the audit managers in the Plausible Third Peer condition assessed that the discount rate should be less than or equal to 11 percent, as compared to 0 percent of the managers in the Two Peer Companies condition (Fisher's Exact Test p-value = 0.044). Thus, replicating Experiment 1 using audit managers, contrast effects still exist for managers.

While comparisons across experiments should be made with caution, we examined if the pattern of the WACC results for audit seniors (Table 2) was different than those of managers (Table 5). A Mantel-Haenszel test indicates significant differences between the audit seniors and the audit managers groups (Mantel-Haenszel $\chi^2 = 14.81$; p = 0.001).[15] A lower percentage of audit managers in the Plausible Third Peer condition (25 percent) agreed with the client-preferred WACC than audit seniors in the Plausible Third Peer condition (64.7 percent) (p = 0.037). Subject to the caveat of making comparisons across experiments, audit managers appear to be less susceptible to contrast effects than audit seniors.

Managers' Documentation

Two authors, blind to the experimental conditions, examined each manager's documentation and coded the responses as either agreeing or disagreeing that the benchmark set suggests that the discount rate is reasonable or made no mention of the set (agreement level = 93.9 percent; Kappa = 0.871, p < 0.001). Panel A of Table 6 presents managers' documentation results for the two experimental conditions. Audit managers' agreement with the client's benchmark set in the Plausible Third Peer

---

[15] A Mantel-Haenszel χ^2 statistic was used to compare across groups because it is effective when some cell counts are zero (Hosmer and Lemeshow 1989; Agresti 1996).

TABLE 5

EXPERIMENT 2
Impact of Benchmark Set Composition on Audit Managers' Discount Rate Assessments

Panel A: Frequency (Percentage) of Audit Managers' Assessments of the WACC by Group[a]

Variable	Two Peer Companies (n = 17)	Plausible Third Peer (n = 16)
Client-Preferred WACC: Less than or Equal to 11%	0 (0.0%)	4 (25.0%)
More Conservative WACC: Greater than 11%	17 (100.0%)	12 (75.0%)
Total	17 (100%)	16 (100%)

Panel B: Test of the Effects of a Plausible Third Peer Company[b]

Comparisons (Audit Managers)	Fisher's Exact Test p-value
Two Peer Companies versus Plausible Third Peer	0.044

[a] Panel A reports the number of audit managers, across conditions, that assessed different WACC rates that should be used for the discounted cash flows. Management stated that they believed an 11 percent WACC was reasonable, while greater than 11 percent WACC was more conservative.
[b] Panel B reports Fisher's Exact Test p-values. The p-value is two-tailed due to non-directional predictions.

condition (25 percent) was higher than managers' agreement in the Two Peer Companies condition (0 percent) (Fisher's Exact Test p-value = 0.043).[16] These results suggest that audit managers' documentation was susceptible to contrast effects.

Impact of Task-Specific Experience on Audit Managers' Assessments

We also examined how two task-specific experience variables impact managers' judgments—familiarity with fair value investments and experience with auditing investments. Prior research indicates that auditors develop more complete knowledge about the task through task-specific experience. Such task-specific experience helps auditors recognize relevant cues in the performance of the task (Bonner 2008; Earley 2002). As managers acquire experience auditing investments and increase their familiarity with fair value judgments, they are likely to develop knowledge structures that enable them to focus on relevant information and ignore irrelevant information. Therefore, task-specific experience may attenuate the impact of the plausible third peer company and reduce contrast effects.

We first examine how audit managers' familiarity with auditing fair value instruments can impact their judgments. Panel A of Table 7 reports the results of a regression analysis of the benchmark condition (Two Peer Companies or Plausible Third Peer), managers' self-reported familiarity with auditing fair value instruments, and the interaction of these two variables on their WACC assessments. The dependent variable was audit managers' assessments of the WACC rate that they believed should be used (*WACC*). We find evidence of skewness in the *WACC* variable. Thus, we rank-transformed the *WACC* in order to avoid estimation problems associated with skewness. Rank transformations can be more effective at avoiding skewness and normality problems as compared to alternative procedures (Conover and Iman 1981). The independent variables are a dummy variable coded 0 = Two Peer Companies condition or 1 = Plausible Third Peer condition (*PEER_COMPANY*); audit managers' ratings of their familiarity with auditing fair value instruments (1 = not familiar at all to 7 = very familiar) (*FV_EXP*); and an interaction term of the two independent variables (*PEER_COMPANY × FV_EXP*).

The regression results are summarized in Panel A of Table 7. The overall regression model is significant ($F_{3,29} = 55.98$; p < 0.001), with an adjusted R^2 of 0.838. With respect to the main effects, *BENCHMARK* is negative and significant (coefficient = −33.65; p = 0.001), indicating that the assessed *WACC* is lower (more client-preferred) in the Plausible Third Peer condition

[16] Given that one of the cells in Panel A has a frequency of 0, Fisher's Exact Test was used to test the comparisons reported in Panel B. A 2 × 2 Fisher's Exact Test was computed, excluding the two managers who provided no mention of the peer companies.

American Accounting Association

TABLE 6

EXPERIMENT 2
Audit Managers' Documentation of Agreement with the Peer Companies

Panel A: Frequency (Percentage) of Auditor Managers Agreeing or Disagreeing that the Peer Company Comparisons were Appropriate[a]

Variable	Two Peer Companies (n = 17)	Plausible Third Peer (n = 16)
Agreed with peer companies	0 (0.0%)	4 (25.0%)
Disagreed with peer companies	16 (94.2%)	11 (68.8%)
No mention of peer companies	1 (5.8%)	1 (6.2%)
Total	17 (100%)	16 (100%)

Panel B: Test of the Effects of a Plausible Third Peer Company[b]

Comparisons (Audit Managers)	Fisher's Exact Test p-value
Two Peer Companies versus Plausible Third Peer	0.043

[a] Panel A reports an analysis of the number of audit managers, across conditions, who agreed or disagreed that the client's peer company comparisons were appropriate.

[b] Given that one of the cells in Panel A has a frequency of 0, Fisher's Exact Test was utilized to test the comparisons reported in Panel B. A 2×2 Fisher's Exact Test was computed, excluding the two managers who provided no mention of the peer companies. The p-value is two-tailed due to non-directional predictions.

relative to the Two Peer Companies condition. We observe that *FV_EXP* is not significant ($p > 0.05$). The interaction term *BENCHMARK $\times$ FV_EXP* is positive, but not significant (coefficient = 2.92; $p = 0.065$). Therefore, fair value familiarity does not attenuate contrast effects.[17]

Panel B of Table 7 examines how audit managers' experience with auditing investments can impact their judgments. We conducted a regression similar to that in Panel A, with the exception of using audit managers' ratings of how often they deal with auditing investments (1 = never to 7 = all the time) as the task experience variable (*INV_EXP*). The overall regression model is significant ($F_{3,29} = 84.41$; $p < 0.001$), with an adjusted R^2 of 0.887. The main effect *BENCHMARK* is negative and significant (coefficient = −45.81; $p = 0.001$), indicating that the assessed *WACC* is lower (more client-preferred) in the Plausible Third Peer condition compared to the Two Peer Companies condition. The coefficient for *INV_EXP* is not significant ($p > 0.05$). However, the interaction term *BENCHMARK $\times$ INV_EXP* is positive and significant (coefficient = 4.71; $p = 0.025$). Thus, as experience with auditing investments increases, the effect of the plausible third peer company decreases given that the assessed *WACC* becomes higher (more conservative) in the Plausible Third Peer condition. It appears that audit managers with more task-specific experience in investing may be less susceptible to contrast effects from the benchmark set composition.[18]

V. CONCLUSION

Our findings suggest that auditors may be inappropriately influenced by the set of peer companies that clients provide to support their estimates or the assumptions underlying these estimates. In Experiment 1, audit seniors assessed a lower WACC

---

[17] As recommended by Aiken and West (1991), as an additional analysis, the continuous independent variable (including the corresponding interaction) was mean-centered to minimize correlations between main effects and interactions. Regressions using the mean-centered variables produced identical results as those reported in Table 7.

[18] We ran the same two regressions as reported in Table 7 using the data from Experiment 1 (audit seniors without structured guidance). The results indicate that the interaction terms in either regression were not significant. This indicates that audit seniors with more task-specific experience may not exhibit the same attenuation of contrast effects as audit managers.

American Accounting Association

TABLE 7

EXPERIMENT 2
Impact of Task-Specific Experience on Audit Managers' Assessments

Panel A: Regression Results—Impact of Familiarity with Auditing Fair Value on Audit Managers' Discount Rate Assessments

$$WACC = b_0 + b_1 PEER\_COMPANY + b_2 FV\_EXP + b_3 PEER\_COMPANY \times FV\_EXP + e$$

Variables	β	Std. Error	t-statistic	p-value[a]
Constant	25.00	7.19	3.47	0.002
Experimental Variables				
PEER_COMPANY	−33.65	9.15	−3.67	0.001
FV_EXP	0.01	1.19	0.01	0.999
PEER_COMPANY × FV_EXP	2.92	1.52	1.96	0.065

Model: n = 33; R^2 = 0.853; Adjusted R^2 = 0.838; F-ratio = 55.98; Significance F < 0.001.

Panel B: Regression Results—Impact of Experience with Auditing Investments on Audit Managers' Discount Rate Assessments

$$WACC = b_0 + b_1 PEER\_COMPANY + b_2 INV\_EXP + b_3 PEER\_COMPANY \times INV\_EXP + e$$

Variables	β	Std. Error	t-statistic	p-value[a]
Constant	16.81	10.16	1.65	0.109
Experimental Variables				
PEER_COMPANY	−45.81	12.61	−3.63	0.001
INV_EXP	1.28	1.59	0.80	0.426
PEER_COMPANY × INV_EXP	4.71	1.98	2.36	0.025

Model: n = 33; R^2 = 0.897; Adjusted R^2 = 0.887; F-ratio = 84.41; Significance F < 0.001.
[a] All p-values represent two-tailed tests due to non-directional predictions.

Variable Definitions:
WACC = audit managers' assessments of the WACC rate that they believed should be used. The variable was rank-transformed in order to avoid estimation problems associated with skewness;
PEER_COMPANY = dummy variable coded 0 = Two Peer Companies condition or 1 = Plausible Third Peer condition;
FV_EXP = audit managers' ratings of their familiarity with auditing fair value instruments (1 = not familiar at all to 7 = very familiar); and
INV_EXP = audit managers' ratings of how often they deal with auditing investments (1 = never to 7 = all the time).

when the client provided a third low similarity peer that made a moderately similar peer with an aggressive WACC appear more appropriate than when the third peer was not provided. Our results also demonstrate that very structured audit guidance for conducting benchmarking analysis may be somewhat successful in reducing, but not eliminating, contrast effects in audit seniors' judgments. Results from Experiment 2 demonstrate that contrast effects are not eliminated when the task is completed by managers. However, it appears that audit managers may be less susceptible to contrast effects from the benchmark set composition as compared to audit seniors. Importantly, as managers' investment experience increases, the impact of a third plausible peer is reduced.

Our results extend our understanding of audit judgment when evaluating estimates. PCAOB inspection reports have noted auditor deficiencies in evaluating the appropriateness of peer companies in benchmarking analysis. This paper identifies a potential cause for this deficiency: contrast effects from the benchmark set composition influence auditors' evaluations of peers. Since auditors are likely to be assessing benchmark data in other fair value areas, this research has implications for many account areas. In addition, the use of benchmark data are ubiquitous in auditing and accounting, including performance evaluation judgments, investor decision making, capital budgeting decisions, and corporate social responsibility evaluations. Therefore, the implications of contrast effects from the benchmark set are broad-ranging.

Our results should be of interest to both practitioners and regulators. Estimates like Level 3 investments require management to provide subjective inputs and assumptions and, in turn, provide support for these inputs. Management has

motivation and opportunity to provide overly persuasive evidence to justify these inputs. Audit firms can make auditors aware of the consequences of such persuasion attempts on the evaluations of estimates. Our results also suggest that if lower-level auditors are provided with very structured guidance to conduct benchmarking analysis, these auditors may be less susceptible to contrast effects than auditors who do not receive such guidance. Equally important, our results suggest that firms should understand the benefits of using more experienced auditors for these subjective tasks.

Our study has limitations that offer opportunities for future research to investigate. Our study considered a small set of attributes for the peer companies for auditors to assess. Future research could determine if some attributes (e.g., industry) are easier to evaluate without comparative data and less susceptible to contrast effects than others. Our study examined a small set of peer companies. Future research could examine how the benchmark set size may impact contrast effects. Future research could also examine the impact of contrast effects in other audit tasks that require the assessment of peer companies. Finally, additional research is needed on audit guidance in the fair value area. The structured guidance that we provided in the first experiment was very task-specific and included step-by-step instructions. As such, it may not be generalizable to other tasks. Importantly, our findings demonstrate that structured guidance and audit experience were only partially successful in reducing contrast effects from the benchmark set composition. Future research should focus on ways to reduce the impact of contrast effects identified in the current study on auditor judgment. Given regulators' concerns over audit quality in this area, ways to improve auditor judgment are warranted.

REFERENCES

Acuitas, Inc. 2017. *Acuitas, Inc.'s survey of fair value audit deficiencies*. Available at: https://www.aicpa.org/content/dam/aicpa/interestareas/forensicandvaluation/newsandpublications/downloadabledocuments/2017-pcaob-fair-value-audit-deficiencies-survey.pdf

Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons.

Aiken, L. S., and S. G. West. 1991. *Multiple Regression: Testing and Interpreting Regressions*. London, U.K.: Sage.

Asay, H. S., T. Brown, M. W. Nelson, and T. J. Wilks. 2017. The effects of out-of-regime guidance on auditor judgments about appropriate application of accounting standards. *Contemporary Accounting Research* 34 (2): 1026–1047. https://doi.org/10.1111/1911-3846.12281

Bhatia, S. 2013. Associations and the accumulation of preference. *Psychological Review* 120 (3): 522–543. https://doi.org/10.1037/a0032457

Bhattacharjee, S., M. J. Maletta, and K. K. Moreno. 2007. The cascading of contrast effects on auditors' judgments in multiple client audit environments. *The Accounting Review* 82 (5): 1097–1117. https://doi.org/10.2308/accr.2007.82.5.1097

Bonner, S. E. 2008. *Judgment and Decision Making in Accounting*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Cannon, N. H., and J. C. Bedard. 2017. Auditing challenging fair value measurements: Evidence from the field. *The Accounting Review* 92 (4): 81–114. https://doi.org/10.2308/accr-51569

Carlson, K., and S. Bond. 2006. Improving preference assessment: Limiting the effect of context through pre-exposure to attribute levels. *Management Science* 52 (3): 410–421. https://doi.org/10.1287/mnsc.1050.0434

Ceres. 2011. A single measure, unbiased results: Ceres, Tellus unveil global initiative for a standardized, comprehensive corporate sustainability rating. (June 9). *CSRwire.com*. Available at: https://www.csrwire.com/press_releases/32411-A-Single-Measure-Unbiased-Results-Ceres-Tellus-Unveil-Global-Initiative-for-a-Standardized-Comprehensive-Corporate-Sustainability-Rating

Christensen, B. E., S. M. Glover, and D. A. Wood. 2012. Extreme estimation uncertainty in fair value estimates: Implications for audit assurance. *Auditing: A Journal of Practice & Theory* 31 (1): 127–146. https://doi.org/10.2308/ajpt-10191

Church, B. K., and L. B. Shefchik. 2012. PCAOB inspections and large accounting firms. *Accounting Horizons* 26 (1): 43–63. https://doi.org/10.2308/acch-50077

Conover, W. J., and R. L. Iman. 1981. Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician* 35 (3): 124–129.

Damisch, L., T. Mussweiler, and H. Plessner. 2006. Olympic medals as fruits of comparison? Assimilation and contrast in sequential performance judgments. *Journal of Experimental Psychology: Applied* 12 (3): 166–178. https://doi.org/10.1037/1076-898X.12.3.166

Dhar, R., and I. Simonson. 2003. The effect of forced choice on choice. *Journal of Marketing Research* 40 (2): 146–160. https://doi.org/10.1509/jmkr.40.2.146.19229

Earley, C. E. 2001. Knowledge acquisition in auditing: Training novice auditors to recognize cue relationships in real estate valuation. *The Accounting Review* 76 (1): 81–97. https://doi.org/10.2308/accr.2001.76.1.81

Earley, C. E. 2002. The differential use of information by experienced and novice auditors in the performance of ill-structured audit tasks. *Contemporary Accounting Research* 19 (4): 595–614. https://doi.org/10.1506/XWDP-PHRH-Q3J9-XLXL

Englich, B., and T. Mussweiler. 2001. Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology* 31 (7): 1535–1551. https://doi.org/10.1111/j.1559-1816.2001.tb02687.x

American Accounting Association

Englich, B., T. Mussweiler, and F. Strack. 2006. Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin* 32 (2): 188–200. https://doi.org/10.1177/0146167205282152

Griffith, E. E., J. S. Hammersley, and K. Kadous. 2015a. Audits of complex estimates as verification of management numbers: How institutional pressures shape practice. *Contemporary Accounting Research* 32 (3): 833–863. https://doi.org/10.1111/1911-3846.12104

Griffith, E. E., J. S. Hammersley, K. Kadous, and D. Young. 2015b. Auditor mindsets and audits of complex estimates. *Journal of Accounting Research* 53 (1): 49–77. https://doi.org/10.1111/1475-679X.12066

Hammersley, J. S. 2006. Pattern identification and industry-specialist auditors. *The Accounting Review* 81 (2): 309–336. https://doi.org/10.2308/accr.2006.81.2.309

Hosmer, D. W., and S. L. Lemeshow. 1989. *Applied Logistic Regression*. New York, NY: John Wiley & Sons.

Huber, J., J. W. Payne, and C. Puto. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research* 9 (1): 90–98. https://doi.org/10.1086/208899

International Forum of Independent Audit Regulators (IFIAR). 2016. *International Forum of Independent Audit Regulators—Report on 2015 survey of inspection findings* (March 3). Available at: https://tinyurl.com/y5zae8yy

International Forum of Independent Audit Regulators (IFIAR). 2017. *International Forum of Independent Audit Regulators—Report on 2016 survey of inspection findings* (March). Available at: https://tinyurl.com/y57dvg89

Low, K-Y. 2004. The effects of industry specialization on audit risk assessments and audit-planning decisions. *The Accounting Review* 79 (1): 201–219. https://doi.org/10.2308/accr.2004.79.1.201

Melin, A., and J. S. Diamond. 2015. How companies justify big pay raises for CEOs. *Bloomberg.com* (June 4). Available at: https://www.bloomberg.com/graphics/2015-executive-pay-peer-groups/

Moran, S., and J. Meyer. 2006. Using context effects to increase a leader's advantage: What set of alternatives should be included in the comparison set? *International Journal of Research in Marketing* 23 (2): 141–154. https://doi.org/10.1016/j.ijresmar.2005.09.009

Noguchi, T., and N. Stewart. 2014. In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition* 132 (1): 44–56. https://doi.org/10.1016/j.cognition.2014.03.006

O'Reilly, D., R. Leitch, and D. Wedell. 2004. The effect of immediate context on auditors' judgments of loan quality. *Auditing: A Journal of Practice & Theory* 23 (1): 89–105. https://doi.org/10.2308/aud.2004.23.1.89

Palmer, J. K., T. J. Maurer, and J. M. Feldman. 2002. Context and prior impression effects on attention, judgment standards, and ratings: Contrast effects revisited. *Journal of Applied Social Psychology* 32 (12): 2575–2597. https://doi.org/10.1111/j.1559-1816.2002.tb02757.x

Public Company Accounting Oversight Board (PCAOB). 2014. *Report on 2013 Inspection of KPMG LLP*. PCAOB Release No. 104-2014-167. Washington, DC: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2016a. *2015 Inspection of MaloneBailey, LLP*. PCAOB Release No. 104-2016-163. Washington, DC: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2016b. *2015 Inspection of PricewaterhouseCoopers LLP*. PCAOB Release No. 104-2016-140. Washington, DC: PCAOB.

Ricchiute, D. N. 1999. The effect of audit seniors' decisions on working paper documentation and on partners' decisions. *Accounting, Organizations and Society* 24 (2): 155–171. https://doi.org/10.1016/S0361-3682(98)00029-4

Shapiro, J. 2010. Benchmarking the benchmarks. *HR Magazine* 4: 43–46.

Shelton, S. W. 1999. The effect of experience on the use of irrelevant evidence in auditor judgment. *The Accounting Review* 74 (2): 217–224. https://doi.org/10.2308/accr.1999.74.2.217

Trueblood, J. S., and J. C. Pettibone. 2017. The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making* 30 (2): 157–167. https://doi.org/10.1002/bdm.1930

Trueblood, J. S., S. D. Brown, A. Heathcote, and J. R. Busemeyer. 2013. Not just for consumers: Context effects are fundamental to decision making. *Psychological Science* 24 (6): 901–908. https://doi.org/10.1177/0956797612464241

Vera-Muñoz, S. C., W. R. Kinney, Jr., and S. E. Bonner. 2001. The effects of domain experience and task presentation format on accountants' information relevance assurance. *The Accounting Review* 76 (3): 405–429. https://doi.org/10.2308/accr.2001.76.3.405

Wong, K. F. E., and J. Y. Y. Kwong. 2005. Comparing two tiny giants or two huge dwarfs? Preference reversals owing to number size framing. *Organizational Behavior and Human Decision Processes* 98 (1): 54–65. https://doi.org/10.1016/j.obhdp.2005.04.002

Yeates, P., P. O'Neill, K. Mann, and K. W. Eva. 2012. Effect of exposure to good versus poor medical trainee performance on attending physician ratings of subsequent performances. *Journal of the American Medical Association* 308 (21): 2226–2232. https://doi.org/10.1001/jama.2012.36515

Zimmerman, J. L. 2017. *Accounting for Decision Making and Control*. 9th Edition. New York, NY: McGraw-Hill Irwin.