Nonprofessional Investor Judgments: Linking Dependent Measures to Constructs

H. Scott Asay
University of Iowa

Jeffrey Hales
The University of Texas at Austin

Cory Hinds
University of Iowa

Kathy Rupar
Georgia Institute of Technology

Acknowledgements: We thank Spencer Anderson, Leo Barcellos, Shana Clor-Proell, Ryan Guggenmos, Pat Hopkins, Lisa Koonce, Kristi Rennekamp, Dan Rimkus, Blake Steenhoven, Rong Su, Sean Wang, Brian White, Michael Wynes, Donnie Young, participants of the Behavioral Accounting Reading Group at the University of Texas at Austin and the doctoral students' reading group at Cornell University. We are also thankful for research assistance from Cassie Mongold, Mandy Ellison, and Peina Liu.

Nonprofessional Investor Judgments: Linking Dependent Measures to Constructs

ABSTRACT

There is limited evidence on the construct validity of the dependent measures commonly used in the literature on nonprofessional investor judgments. In this paper, we first survey the literature to understand the types of dependent measures typically used by researchers. We then conduct factor analyses to uncover linkages between dependent measures and the constructs underlying these nonprofessional investor judgments. Our results suggest that, while the wide variety of dependent measures can appear on their face to represent many nuanced economic constructs, these measures load onto three distinct factors. These factors relate to nonprofessional investors' (1) expectations regarding future firm performance and value, (2) holistic perceptions of the firm, and (3) evaluations of the risk associated with investing in the firm. Next, we provide recommendations for selecting, analyzing, and reporting dependent measures in future research. Finally, we conclude by providing directions for future research to further our understanding of the judgments made by investors.

Keywords: investor judgments, construct validity, experimental design

I. INTRODUCTION

The literature on individual investors' judgments has grown substantially since its inception in the 1990s. Studies within this literature have examined a wide variety of research questions with important implications, including how investors respond to information presentation (e.g., Hopkins 1996; Maines and McDaniel 2000), disclosure language (e.g., Hales, Kuang, and Venkataraman 2011; Rennekamp 2012; Tan, Wang, and Zhou 2014; Elliott, Rennekamp, and White 2015; Asay, Elliott, and Rennekamp 2017), earnings characteristics (e.g., Elliott 2006; Erickson, Hewitt, and Maines 2017; Koonce and Lipe 2017), disclosure venue (e.g., Elliott, Hodge, and Sedor 2012; Cade 2018; Elliott, Grant, and Hodge 2018), and sustainability information (Elliott, Jackson, Peecher, and White 2014, Guiral, Moon, Tan, and Yu 2020; Johnson, Theis, Vitalis, and Young 2020). Over the years, researchers have periodically written review articles to summarize the major areas of research within this literature with an emphasis on understanding the independent variables that influence investor judgment (e.g., Libby, Bloomfield, and Nelson 2002; Libby and Emett 2014). In this paper, we turn our attention to the other side of the equation – namely, the ***construct validity*** of the dependent measures used to represent nonprofessional investor judgments in this literature.[1]

From the perspective of scientific discovery, increasing the construct validity of the dependent measures within the investor judgment literature would improve our ability to compare studies within the literature, link important findings to the archival literature, and generalize findings to real-world settings. From the perspective of individual researchers, a clearer understanding of the validity of various measures would help us more effectively

[1] By construct validity, we mean the extent to which operational measures map onto the constructs they are intended to represent (Shadish, Cook, and Campbell 2002). We provide a glossary of key terms in Appendix A. To facilitate use of the glossary, we emphasize in ***bold italics*** defined terms the first time they are used in the text.

operationalize specific dependent constructs of interest, select dependent measures that maximize experimental power for detecting meaningful effects, articulate our choices in manuscripts and during the peer-review process, and facilitate testing more descriptive models of investor judgments. Given these potential benefits, the construct validity of the dependent measures used in these studies merits explicit attention.[2] Our goals in this paper are threefold: (1) improve our understanding of the constructs captured by investor-oriented dependent measures, (2) provide future researchers with a useful resource when selecting (and later defending) their dependent measure choices, and (3) stimulate future research on investor judgments.

We first seek to understand the extent to which various dependent measures are used in the literature on investors judgments.[3] To do so, we identify research relating to individual investors published between 1990 and the first quarter of 2020 within six top accounting journals, and we catalogue the dependent measures collected in each study. Through this process, we identify 90 articles and their dependent measures. We group the dependent measures with similar wording and find that researchers have generally relied on eight types of judgments: firm value, expected future earnings, expected future cash flows, firm risk, earnings multiples, investment attractiveness, investment likelihood, and investment amount. These different types of judgments are elicited using dependent measures that are both qualitative and quantitative in nature, and they are elicited from a variety of sample pools (e.g., general public, students, and experienced professionals). This survey of the literature highlights that researchers have used

[2] As noted by Cook and Campbell (1979), "most applied experimental research is much more oriented toward high construct validity of effects than of causes. This is entirely understandable, for what one wants to see is evidence that the social problem being addressed is at least partially ameliorated – not any problem, but *the* major problem as generally conceived. Thus, great care goes into measuring outcomes, for unless a rigorous measure…is used which most competent persons believe to be reasonable, the research is likely to be seen as 'irrelevant'" (p. 63-64).
[3] Most JDM studies present equity investors with dependent measures framed as judgments or hypothetical decisions. For brevity, we refer to these collectively as investor judgments and limit our focus to these dependent measures.

many different measures (at an operational level) to draw conclusions about many different types of investor judgments.

Having shed light on the range of dependent measures historically used in the literature and the types of investor judgments researchers have examined, we next turn to an analysis of the extent to which these measures capture dimensions of distinct or related constructs. While the judgment categories we identify may be conceptually unique, they may be operationally redundant or capture different dimensions of the same underlying construct related to investor judgment. We use a two-stage empirical analysis to provide evidence of the ***structural validity*** and ***dimensionality*** of the dependent measures used in prior work and link these measures to the constructs they most closely represent. In doing so, we limit our analysis to the judgments of nonprofessional investors proxied by Amazon Mechanical Turk workers.

In the first stage, we use a $2 \times 2 \times 2$ experimental design to create eight vignettes, each of which provides a brief description of a company. These descriptions vary in terms of both financial and nonfinancial performance. Our manipulations were informed by valuation frameworks that contain numerator effects (such as expected future earnings) and denominator effects (such as investment risk). In addition, we also manipulate nonpecuniary factors (such as details about the company's business practices) that may affect investors' perceptions of the company. Each participant is randomly assigned to evaluate a single firm vignette using 38 dependent measures derived from our survey of the literature. With this data, we then use ***exploratory factor analysis*** (EFA) to examine the variance-covariance structure of these measures. Additionally, we specify a ***measurement model*** to evaluate the link between our experimental manipulations and the factors uncovered by the EFA.

Results from the EFA reveal that the dependent measures derived from the literature load onto three distinct constructs. The first factor is largely comprised of measures that relate to nonprofessional investors' expectations regarding future firm performance and value. The second factor is largely comprised of measures that relate to nonprofessional investors' holistic perceptions of the firm (where these perceptions could reflect pecuniary and/or nonpecuniary aspects of corporate performance). The third factor is largely comprised of measures that relate to nonprofessional investors' evaluations of the risk associated with investing in the firm. As part of the construct validation process, we also test a measurement model that indicates that the first factor is most sensitive to our manipulation of expected future earnings, the second factor is most sensitive to our manipulation of the company's business practices, and the third factor is most sensitive to our manipulation of investment risk.

In the second stage, we further verify the robustness of the factor structure identified in the first stage to different research settings. To do so, we reexamined the studies we used to identify dependent measures to see which articles include either a full or partial set of experimental materials. From the 27 articles that include an instrument for at least one condition, we create a total of 93 vignettes. These vignettes offer substantially more variation in the information provided to participants than the eight firm vignettes we used in the first stage of our empirical analysis, but also in a way that is less controlled. As before, each experimental participant is randomly assigned to read one of these vignettes and then evaluate that firm using the same set of dependent measures as in our first stage. We use the responses to perform a ***confirmatory factor analysis*** (CFA) using the EFA ***factor loadings*** to assign the dependent measures to one of three factors. The CFA results indicate that the three-factor measurement model fits the data well.

Taken together, the findings from our EFA and CFA can inform researchers as they select, analyze, and report dependent measures. Going forward, we recommend the following five-step process for researchers: First, drawing on theory, specify a dependent construct. Our analyses point to three judgment-related constructs that may be of interest to researchers studying investor judgment. Second, create a ***measurement scale*** by selecting a set of dependent measures that are expected to capture different dimensions of the construct of interest. Our findings can inform this process for researchers interested in the three constructs we identify. Third, provide evidence of the unidimensionality of the dependent measures selected in step two. Fourth, report descriptive statistics for each dependent measure by experimental condition. Fifth, use the arithmetic mean to combine the dependent measures to test hypotheses.

Following this five-step process yields several advantages. By selecting a set of dependent measures related to a specified construct, researchers can capture different dimensions of that construct, increase measurement scale reliability, and report results without introducing concerns about researcher degrees of freedom. This five-step process should also improve the consistency in conducting and reporting analyses by providing guidance on which dependent measures are most likely to consistently load onto the same factor and on which dependent measures to analyze separately versus jointly. In addition, this process constrains selective reporting of dependent measures, increasing the reliability and credibility of the associated findings. Finally, selecting a theory-driven construct and data-driven measures of that construct should increase statistical power.

The remainder of this paper proceeds as follows. In Section II, we describe our survey of the dependent measures used in the literature and report a descriptive analysis of these measures. In Section III, we report the results of our factor analyses. In Section IV, we recommend a five-step process for selecting, analyzing, and reporting dependent measures. In Section V, we highlight potential avenues for future research. Section VI concludes the paper.

II. SURVEY OF THE LITERATURE AND DESCRIPTIVE ANALYSIS

To link investor judgment dependent measures to the constructs they represent, we begin by surveying the experimental financial research literature to identify and catalogue the dependent measures used in prior work. We then identify potential underlying constructs associated with these dependent measures by grouping them together based on similarities in their wording (i.e., use of terms like firm value, earnings, cash flow, risk, price-earnings multiples, investment attractiveness, likelihood of investment, and investment amount). We also provide descriptive evidence on the frequency with which different categories of dependent measures are used. We discuss each of these steps in more detail below.

Survey of the Literature

Identification of Studies

To identify key experimental studies on investor judgments in accounting, we focused on six accounting journals: Accounting, Organizations and Society (AOS), Contemporary Accounting Research (CAR), Journal of Accounting and Economics (JAE), Journal of Accounting Research (JAR), Review of Accounting Studies (RAST), and The Accounting Review (TAR). We then used the Brigham Young University Research Rankings database (Summers and Wood 2018) to identify all papers published in these journals between the years 1990 and 2018 that were classified as both financial and experimental.[4] We next reviewed these

[4] The data was retrieved in January 2019. We began our search in 1990 because, as described by Libby et al. (2002), several market inefficiencies were documented around this time (for reviews, see Fama 1998; Thaler 1999; Kothari 2001), paving the way for a subsequent resurgence of experimental research on individual investors. The earliest paper meeting our selection criteria is Hirst, Koonce, and Simko (1995).

papers to ensure that each study is primarily an investor judgment and decision-making (JDM) study. For completeness, we include studies using professional analysts for participants, as analysts also make investment-related judgments. While our focus remains on nonprofessional investors, these studies are informative to our goal of better understanding the dependent measures used to study investor judgments.[5] By way of contrast, our sample of studies excludes studies where participants acted in the role of a manager or an auditor. We also exclude studies examining assessments of creditworthiness for practical, rather than conceptual, reasons because the vast majority of studies have focused on investments in equity securities. Our final scoping decision is to exclude studies using experimental markets because experimental markets studies face a different set of design issues than studies from the JDM paradigm.[6]

To confirm the completeness of the identified papers, we also conducted an independent search for papers published between January 1990 and March 2020 that investigate investor valuation judgments using experimental methods. First, we searched for papers published between January 1990 and June 2019 for papers published within the selected journals that contained any of the following phrases: investor judgment, investor valuation, investor decision, or investor judgments. We then reviewed the resulting papers for relevance as described above. Second, we performed a forward search of the papers identified to look for any related papers that also fit our criteria as described above and then conducted a hand review of all 2019 and quarter one 2020 volumes of the six journals identified above. Our final set of papers includes 90 articles (see Appendix B).

Identification of Dependent Measures

[5] Including these studies also helps us to generate an overinclusive set of dependent measures for our later analyses, consistent with best practices for construct identification and validation (DeVellis 2016).
[6] See Libby et al. (2002) for a discussion of the benefits and costs associated with laboratory markets.

Having constructed a sample of JDM studies of equity investor judgments, we next examined each paper within our sample to identify the dependent measures used in the literature. For each paper, we recorded the primary dependent measure(s) used in each experiment. Because we are interested in the key constructs of interest in this literature, we restrict our focus to primary dependent measures and, therefore, exclude measures of second-order effects (e.g., confidence in participants' judgments) and measures that are commonly used as process measures (e.g., management credibility).[7]

Descriptive Analysis

Grouping of Dependent Measures

Next, we grouped the dependent measures from the papers in our sample based on the similarity of the words used in the phrasing of the measures. This process resulted in the following groups: firm value, expected future earnings, expected future cash flows, firm risk, earnings multiples, investment attractiveness, investment likelihood, and investment amount. Figure 1 depicts the relative frequency with which researchers selected measures from each category, and Figure 2 depicts the variation in the number of dependent measures elicited within a given study.

[INSERT FIGURE 1]

[INSERT FIGURE 2]

Preliminary Assessment of Construct Validity

Our descriptive analysis and categorization is an important first step in identifying

[7] One study, Mercer (2004), uses measures that capture management credibility as the primary dependent variables of interest. Management credibility is a multi-dimensional construct and is an important construct in its own right (see Mercer 2005). However, because credibility is more commonly used as a process variable, we view this construct as outside the scope of our study.

investor-oriented dependent measures and potential corresponding theoretical constructs. This process is commonly referred to as the substantive phase of construct validation (Loevinger 1957; Flake, Pek, and Hehman 2017). Our categorization highlights that different dependent measures may capture unique types of judgments made by investors. However, little empirical evidence exists to inform researchers regarding the extent to which different dependent measures can differentiate between various constructs. Further, our categorization is necessarily subjective and only offers ***face (or translation) validity***. However, face validity provides very limited evidence of construct validity. For example, the assumption of face validity can be wrong in the sense that the link between the operational measure and intended theoretical construct may be weak or non-existent. In addition, the community of accounting academics may disagree about the extent to which face validity is evident. In the case of experimental accounting research, relying on face validity has resulted in a host of dependent measures with little guidance on how to select, analyze, and report those measures. This limits the comparability of results across studies and could potentially lead to erroneous inferences.

To move beyond a reliance on face validity, we next test the psychometric properties of the measures using more rigorous, empirical approaches. Specifically, we use an EFA to determine the underlying factor structure of the dependent measures commonly used in the investor judgment literature and a CFA to confirm the validity of the factor structure identified by the EFA.[8] For these analyses, we focus on nonprofessional investors and dependent measures that are qualitative in nature. We discuss opportunities for future research to test the generalizability of our findings to both qualitative and quantitative measures and across different samples of investors (both nonprofessional and professional) in Section V.

[8] We obtained Institutional Review Board approval for the studies in this paper.

III. EXPLORATORY AND CONFIRMATORY FACTOR ANALYSIS

Exploratory Factor Analysis

Background and Procedures

The basic assumption of factor analysis is that, for a collection of observed measures, there exists either a single underlying ***latent construct*** or a set of underlying latent constructs that can explain the interrelationships among those measures. To identify these underlying latent constructs (or factors), an EFA first partitions the total variance of the variables into two subcomponents: ***common variance*** and ***unique variance***. Common variance is the amount of total variance that is shared among the set of measures, while unique variance is the complement of common variance and represents variance that is either specific to a particular measure or results from measurement error. In an EFA, measures are allowed to freely cluster together based on their shared variance. This allows the data to speak for itself in identifying factors that explain the common variance among measures and identifying which measures best represent those factors.

To generate a sample of dependent measures to use in our analysis, we begin with the dependent measures identified from prior literature and combine measures that are nearly identical or redundant. In addition, we include several measures that elicit more general perceptions about the firm.[9] To minimize response variance attributable to methodological artifact, we adapt all scales to be 7-point Likert-style response scales and use qualitative labels

[9] Consistent with best practices for construct identification and validation, our goal was to include a large, overinclusive set of measures (DeVellis 2016). In choosing to add measures that elicit nonprofessional investors' general perceptions, we were guided by research on attribute substitution, which indicates that individuals faced with a complex assessment sometimes substitute that assessment for an easier one (Kahneman and Frederick 2002). Results are unchanged if we exclude the general perception measures from the analysis.

for all measures (e.g., Very low to Very high; Very weak to Very strong; etc.). The use of qualitative scales also aligns with prior work suggesting that they are better suited for dependent measures elicited from nonprofessional investors – the participant group we recruited for our experiment (Frederickson and Miller 2004). These procedures result in 38 dependent measures that we use in the EFA. Each measure and its associated scale are included in Table 1.

[INSERT TABLE 1]

Psychology and other social sciences often use EFA to uncover latent constructs that describe individuals' stable traits and attitudes. In contrast, financial accounting experiments typically involve vignettes that include manipulations of key institutional features of an accounting setting (Libby et al. 2002). In designing a set of such vignettes, our goal was to create variation across a variety of firm attributes that the dependent measures in the literature appear to capture (on their face). Our choice of manipulations across the series of vignettes was informed by valuation frameworks that contain numerator effects (such as expected future earnings) and denominator effects (such as investment risk). In addition, we also manipulate nonpecuniary factors (such as details of a company's business practices) that may affect investors' perceptions of the company.

More specifically, we use a 2 × 2 × 2 experimental design in which we manipulate a firm's expected future earnings (high vs. low), investment risk (high vs low), and environmental, social, and governance (ESG) performance (high vs. low). The full-factorial combination of these firm attributes results in eight unique firms. Combined with variation in the attributes and preferences of nonprofessional investors, this approach ensures substantial variation in the theoretical constructs that the set of dependent measures might plausibly represent and allows us to uncover the underlying factor structure captured by the set of dependent measures.

Importantly, the primary goal of the independent variable manipulations is to create meaningful variation in responses that allows the EFA to detect which dependent measures co-vary and which dependent measures capture distinct constructs. As we discuss later, a secondary goal of the independent variable manipulations is to help interpret the results of the EFA.

We manipulate the firm's expected future earnings by telling participants in the high (low) condition that the company's managers forecast earnings growth to be 15% (1%) next year and that the average of the individual analyst forecasts also projects strong (weak) earnings growth over the next several years. We manipulate the firm's investment risk by telling participants in the high (low) condition that growth in demand for the company's products is relatively volatile (stable), causing future earnings to be somewhat unpredictable (fairly predictable). Additionally, we tell participants that individual analysts' earnings growth forecasts are considerably different (generally similar), with some significantly (slightly) lower and some significantly (slightly) higher than the average forecast of earnings growth. Finally, we manipulate details of the firm's business practices by telling participants in the high (low) ESG performance condition that the company is widely viewed as having a positive (negative) impact on the environment and is known for treating its employees well (poorly). Additionally, we tell participants that recent news coverage has praised (criticized) the company for supporting (damaging) small communities and opposing (paying) bribes to foreign governments. Each independent variable is presented as a separate bullet point, with the order of presentation randomly determined (see Appendix C).

Participants

Given our focus on nonprofessional investors, we recruited participants from Amazon Mechanical Turk. Using this recruiting platform allows us to obtain a large sample size at a relatively low cost. In addition, prior research suggests online workers can be a suitable recruitment pool for the accounting settings we use to validate the measures (Ferrell, Grenier, and Leiby 2017). However, we also collected information about the participants' investing experience.[10] Collecting this type of demographic information allows us to later test whether our results are robust to different characterizations of a nonprofessional investor.

Participants were randomly assigned to one of the eight experimental conditions and asked to respond to all 38 dependent measures.[11] The order of the dependent measures was fully randomized between participants. Each dependent measure appeared on a separate screen along with the firm information for their assigned vignette, which was displayed throughout the study. Our final sample includes responses from 999 participants, 52 percent of whom are male and who have a mean (median) age of 40 (38).[12] Further information about participant demographics can be found in Table 2.

[INSERT TABLE 2]

Results and Discussion

Table 3 presents the correlations among the dependent measures. The correlation matrix serves as the input that the EFA uses to determine the underlying factor structure. From the correlation matrix, eigenvalues for each factor can be computed to determine the relative importance of each factor in explaining the underlying data. We use parallel analysis to determine the number of factors with significant eigenvalues (Horn 1965). This approach

---

[10] We collected information on participant demographics and investing experience in a prescreening survey. In doing so, we followed best practices suggested by Bentley (2021).
[11] The mean (median) time that it took participants to complete the EFA was 10.8 (8.4) minutes. We paid participants who completed the entire study $2, an average effective hourly rate of approximately $11.
[12] We collected 1,027 responses from Amazon Mechanical Turk participants. From these responses, 26 were removed for not completing the entire survey. An additional two responses were not included in our final sample because the responses came from a duplicate IP address.

compares the observed eigenvalues to eigenvalues derived from a Monte-Carlo simulated matrix with random data and the same numbers of observations and measures as the original data. Factors derived from the original data are retained when their eigenvalues are greater than the eigenvalues derived from the simulated data. Smaller eigenvalues are attributed to random noise. The result of the parallel analysis suggests that the appropriate number of factors to extract is three.[13] Based on these results, we conduct an EFA to extract three factors. We use a maximum likelihood estimation method and a direct oblimin oblique rotation. However, the factor solution is robust to other estimation methods and rotations.

[INSERT TABLE 3]

The EFA results are presented in Table 4, Panel A. The three-factor model explains 77 percent of the total variance in the 38 dependent measures, several measures load on each factor, and there are few measures that load on multiple factors (with a factor loading cutoff of 0.30 – see Comrey and Lee 1992).[14] Most items have high ***common variance*** or ***communality***, indicating that there is considerable overlap among the measures that load on the same factor. The ***item complexity*** score indicates the degree to which an item (or measure) reflects a single construct. The score will equal 1 if an item loads on exactly one factor, 2 if it evenly loads on two factors, and so on. The mean item complexity score from the EFA is 1.1.

[INSERT TABLE 4]

---

[13] Empirical evidence suggests that parallel analysis is the most accurate method for determining the number of factors (Zwick and Velicer 1986; Velicer, Eaton, and Fava 2000). Alternative approaches include the eigenvalue greater-than-one rule (Kaiser 1960) and the scree plot test (Cattell 1966). These alternative methods also suggest extracting three factors.

[14] We note that DM17, DM18, and DM19 cross-loaded on Factor 1 and Factor 2 while measures that appear similar on their face (e.g., DM15, DM16, DM20, DM21, and DM22) loaded only on Factor 1. Several factors can cause differences in how measures are interpreted and evaluated by participants, such as measure wording, response scale labels, and specifics about the experimental instrument and independent variables. We hesitate to speculate as to why the factor structure emerged as it did with respect to specific dependent measures.

Next, we verify the robustness of the factor solution by conducting EFAs with alternative specifications of nonprofessional investors in our sample.[15] First, we restrict our sample to the 529 participants who meet the Center for Audit Quality's (CAQ) definition of an individual investor, which is an individual that (1) is an adult, (2) is the primary decision-maker of their household, or shares that responsibility equally with another household member, and (3) has $10,000 or more in investments, including stocks, bonds, mutual funds, IRAs, 401(k) plans, and the like (CAQ 2019). Second, we restrict our sample to the 567 participants who indicate that they have invested or have plans to invest in individual company stocks, consistent with prior research that often screens on or reports this type of information for Amazon Mechanical Turk participants (e.g., Rennekamp 2012; Koonce, Miller, and Winchel 2015; Asay et al. 2017; Kelton and Montague 2018; Cardinaels, Hollander, and White 2019). Third, we restrict our sample to the 907 participants who indicate they have invested, or have plans to invest, in any of the following asset types: individual stocks, mutual/index funds, 401(k), or government/corporate bonds. Finally, to try to get at actual investors as opposed to those who merely plan to invest, we restrict our sample to the 379 participants who indicate that they have invested or have plans to invest in individual company stocks only and indicate that they have at least $10,000 invested across all asset classes.

In untabulated analyses, we find that for each of the above alternative specifications for identifying nonprofessional investors, the dependent measures load on all the same factors and the same four measures cross-load on multiple factors in the same manner as the original

[15] We use responses to the following three questions to identify participants with varying levels of investing experience: (1) Indicate whether you have ever invested, or plan to invest in the future, in any of the following (check all that apply): individual company stocks, mutual funds, index funds, 401(k), government bonds, corporate bonds; (2) Are you the primary decision-maker regarding your household's savings and investments, or do you share this role equally with another household member; (3) Indicate the approximate total value of your investments in stocks, bonds, mutual funds, IRAs, 401(k) plans, and the like.

analysis. Additionally, the correlation between factors and the amount of variance explained by the three factors is qualitatively similar as the previously reported EFA results. Note that the robustness of our results to these different specifications is not an indication that the definitions are interchangeable. Instead, the robustness tests only provide evidence that the factor solutions we elicit using an EFA on the 38 dependent measures are similar when we use any of these three subsets of the entire sample. Importantly, these results do not suggest that Amazon Mechanical Turk workers are suitable participants for all nonprofessional investor studies. That question will be informed by other factors, such as the type of independent variables being manipulated, the setting type, and the task complexity. We refer researchers to the excellent guidance provided in prior work for more discussion on how to select appropriate participants when considering these other design choices (Elliott, Hodge, Kennedy, and Pronk 2007; Farrell et al. 2017; Krische 2019; Libby et al. 2002).

With the factor solution in hand, the dependent measures that load on each factor provide some insight into the theoretical construct each factor may represent (see Table 5). Measures that load on the first factor relate to nonprofessional investors' expectations regarding future firm performance and value (i.e., expected future performance and value). They include earnings and cash flow forecasts, growth expectations, and stock valuations. Measures that load on the second factor relate to nonprofessional investors' holistic perceptions of the firm. They include a more diverse set of measures, some focusing on general perceptions and the favorability of the company's stock as an investment, and others focusing on changes in one's buy, sell, or hold position in the company.[16] We note that there is a high correlation between Factor 1 and Factor 2

[16] Interestingly, the measures that are framed as hypothetical investment decisions (e.g., DM24, DM25, DM26, DM28, and DM30) do not appear to behave differently from measures that are framed as judgments. Despite appearing on their face to reflect decisions or choices (i.e., face validity), their statistical properties are not different

suggesting that the holistic perceptions of the firm captured by Factor 2 may comprise some of the aspects of expected future performance and value (i.e., Factor 1) as well as some potentially nonpecuniary components impacting investors' perceptions about a company (e.g., alignment with personal values or other hedonic aspects to investing). We further discuss this correlation in Section V when offering suggestions for future research. Measures that load on the third factor relate to nonprofessional investors' evaluations of the risk associated with investing in the firm (i.e., investment risk). They include evaluations of overall investment risk as well as expectations of stock price decline.

[INSERT TABLE 5]

While our approach is empirically driven, the interpretation of the factors can be subjective. To better link the dependent measures to the theoretical constructs, we specify a measurement model (see Table 4, Panel B) that allows us to evaluate the influence our three independent variables have on each factor. To do so, after excluding the four dependent measures that cross-load onto multiple factors in the EFA, we extract Factors 1, 2, and 3 from the dependent measures that loaded on each factor. Next, our model runs three simultaneous regressions of Factor 1, Factor 2, and Factor 3 on the three independent variables (expected future earnings, investment risk, and business practices). Last, we allow for each of the three factors to co-vary with one another. The model is similar in spirit to a $2 \times 2 \times 2$ MANOVA with Factor 1, Factor 2, and Factor 3 as the dependent variables, and each factor measured by the respective dependent measures that loaded on the factors from the EFA.

As reported in Table 4, Panel C, we find that the manipulation of future expected

than the general perception measures. As we discuss in Section V, we call for future research to improve our understanding of the investor judgment and decision-making process, including understanding how the constructs we identify translate into actual decisions by investors.

earnings is most strongly associated with Factor 1, the manipulation of business practices is most strongly associated with Factor 2, and the manipulation of investment risk is mostly strongly associated with the Factor 3. This provides confidence that our interpretation of the factor solution and associated underlying constructs is appropriate.

Confirmatory Factor Analysis

Background and Procedures

CFA is a form of structural equation modeling that is used to test or "confirm" the goodness of fit of a hypothesized factor structure underlying a set of measures. In a CFA, a model is specified based on theory or a prior analytical result. In general, a CFA can complement an EFA in at least two important ways. First, using a new sample, CFA can be used to confirm that the factor structure revealed through an EFA was not a result of chance, and model fit statistics can be calculated to determine whether the model is a good fit for the data. Second, the hypothesized model can be compared against alternative model specifications.

In our specific case, the EFA revealed a three-factor model with the following identified constructs: (1) expected future performance and value, (2) holistic perceptions of the firm, and (3) investment risk. The primary advantage of the EFA design is that it ensured relatively powerful manipulations of theoretically important constructs (Libby et al. 2002). However, a potential concern with this design is that the observed three-factor model emerged as an artifact of the salient operationalization of the three manipulations.

In our CFA, we test the robustness of the three-factor model using experimental materials from published studies investigating investor judgments. Starting with the 90 articles used in our descriptive analysis, we identify 32 papers that published either full or partial experimental materials. Using the experimental materials from these papers, we create a set of 93 experimental conditions for use in our CFA.[17] The rest of the experimental procedures are the same as those outlined for the EFA, except that the materials from the 93 published experimental conditions replace the eight vignettes we constructed for the EFA.

The 93 experimental conditions we extracted should be broadly reflective of the type of investment scenarios studied in the investor JDM literature. They include manipulations of a variety of important institutional accounting variables of interest, such as earnings guidance consistency, CSR/ESG performance, readability and other narrative disclosure features, information presentation and saliency, cautionary disclaimers, the frequency and magnitude of earnings beats, and the sign of firm performance. The materials also include several different disclosure mediums, including financial statements, footnote disclosures, periodic earnings announcement press releases, auditor reports, analyst reports, and social media feeds.

As was the case with our EFA, the independent variable manipulations are not of primary interest, *per se*. Rather, we are interested in creating variation in firm characteristics and attributes that might meaningfully influence investors' perceptions. This allows us to test empirically how well the factor structure identified by the EFA explains variation in participants' judgments, as captured by their responses to the dependent measures. As a result, the CFA can increase confidence that the factor structure we identify is reasonably generalizable for nonprofessional investors (as proxied by Amazon Mechanical Turk workers) across a wide variety of important accounting settings.

Participants

[17] The 32 papers had published materials resulting in 129 experimental conditions. We eliminated materials from five papers because they were too long to reasonably present to participants on a single page. Next, some papers included materials for a greater number of conditions than others, with some providing materials for just one condition and others providing materials for up to 14 conditions. To avoid over-representing any one paper, we select up to four experimental conditions from each paper. For papers with more than four conditions, we select conditions randomly. Our final sample includes 93 experimental conditions from 27 published papers.

We recruited an independent sample of participants (who did not participate in the EFA) from Amazon Mechanical Turk, and we randomly assigned each participant to one of the 93 conditions. Each participant responded to the same 38 dependent measures included in the EFA and we again fully randomize the order in which they answered the dependent measures.[18] Our final sample includes responses from 998 participants.[19] Approximately 52 percent of participants in the sample are male and the mean (median) age is 40 (37). Further information about participant demographics can be found in Table 6.

[INSERT TABLE 6]

Results and Discussion

To conduct a CFA, we specify and then test a measurement model like the one identified in the stage one EFA. After excluding the four dependent measures that cross-loaded onto multiple factors in the EFA, we use the remaining 34 dependent measures to develop three factors based on the factor loadings from the EFA. For example, we assign the 17 dependent measures that load on Factor 1 in the EFA to a single factor labeled "Factor 1" in the CFA measurement model. Next, because the factors are not expected to be orthogonal, we allow the three factors to co-vary with one another. The CFA model and factor loadings are reported in Table 7, Panel A.

[18] The mean (median) time that it took participants to complete the CFA was 11.0 (8.1) minutes. We paid participants who completed the entire study $2, an average effective hourly rate of approximately $11.

[19] We collected 1,099 responses from Amazon Mechanical Turk participants. From these responses, 99 were removed for not completing the entire survey. An additional two responses were not included in our final sample because the responses came from a duplicate IP address. After removing these 101 responses, our final sample consists of 998 responses. In determining the sufficiency of this sample size, we rely on recommendations from prior literature. Most recommendations come in terms of either total N or the minimum ratio of N to the number of measures being analyzed. In terms of total N, recommendations have ranged from a minimum of 100 (Kline 1979; Gorsuch 1983), 200 (Guilford 1954), 250 (Cattell 1978), or upwards of 500 or 1,000 (Comrey and Lee 1992). With respect of the ratio of N to the number of measures analyzed, recommended ratios have varied from 3 to 6 (Cattell 1978), 5 (Gorsuch 1983), and 10 (Everitt 1975). Our sample includes 998 participants, and the ratio of participants to measures analyzed is greater than 25 (998:38).

In Panel B, we present fit statistics for five goodness-of-fit tests: the ***model chi-square***, the ***Comparative Fit Index (CFI)***, the ***Tucker-Lewis Index (TLI)***, the ***Root Mean Square Error of Approximation (RMSEA)***, and the ***Standardized Root Mean Square Residual (SRMR)***. Overall, the model fit is good, with χ^2 = 2,359.67 (df = 524; p < 0.001), CFI = 0.958, TLI = 0.955, RMSEA = 0.059 (90% confidence interval = [0.057, 0.062], p for H_0: RMSEA ≤ 0.05 is less than 0.001), SRMR = 0.033. While goodness-of-fit tests are rarely definitive, the results of these tests provide preliminary evidence that the dependent measures we extracted from the literature are well captured by our three-factor model.[20] In addition, as was the case with the EFA, we draw similar conclusions when repeating these tests using various partitions of our sample to identify nonprofessional investors.[21]

[INSERT TABLE 7]

Next, in untabulated analyses, we examine two alternative model specifications. First, we examine a one-factor model by combining all 34 dependent measures into a single factor, which would suggest that nonprofessional investors' responses to the dependent measures cannot meaningfully differentiate between different types of judgments. We find that a one-factor model does not fit the data well, with χ^2 = 4,980.33 (df = 527; p < 0.001), CFI = 0.898, TLI = 0.891, RMSEA = 0.092 (90% confidence interval = [0.090, 0.094], p for H_0: RMSEA ≤ 0.05 is less than

[20] χ^2 tests the null hypothesis that model fit is perfect. However, testing that the model fit is perfect is usually too conservative (particularly with large sample sizes), so alternative measures of model fit are often used in conjunction with χ^2. CFI and TLI ≥ 0.95 indicates good model fit, and CFI and TLI ≥ 0.90 indicates acceptable model fit. RMSEA ≤ 0.05 indicates good model fit, and RMSEA ≤ 0.08 indicates acceptable model fit. SRMR ≤ 0.06 indicates good model fit, and SRMR ≤ 0.08 indicates acceptable model fit. Indices that fall outside the specified cutoffs are generally considered to indicate poor fit. See Hu and Bentler (1999) for descriptions of model fit indices and justification for prescribed cutoffs.

[21] We test the CFA results on the same four alternative specifications for identifying nonprofessional investors in our sample as we did with the EFA in Section III. Those four subsamples, when applied to this participant pool, resulted in 558, 607, 912, and 414 participants, respectively. All reported CFA results are robust to the alternative specifications.

0.001), SRMR = 0.049. Next, as reported in Table 4, Panel A, our EFA revealed that Factor 1 and Factor 2 are positively correlated ($\rho = 0.78$) and that both are negatively correlated with Factor 3 ($\rho_{1,3} = -0.45$; $\rho_{2,3} = -0.49$).[22] Therefore, we examine a two-factor model where the 17 dependent measures that load on Factor 1 and the 12 dependent measures that load on Factor 2 are combined into a single factor. This two-factor model provides acceptable model fit, with χ^2 = 3,511.73 (df = 526; $p < 0.001$), CFI = 0.932, TLI = 0.927, H_0: RMSEA = 0.075 (90% confidence interval = [0.073, 0.078], p for H_0: RMSEA $\leq$ 0.05 is less than 0.001), SRMR = 0.035. Finally, the three-factor model (which is nested within both the one-factor and two-factor models) improves model fit over the one-factor model (χ^2 Difference = 2,620.7; df difference = 3; p < 0.001) and two-factor model (χ^2 Difference = 1,152.1; df difference = 2; p = < 0.001).

In summary, the three-factor model provides considerable improvement in model fit over the one-factor and two factor model, with only the three-factor model achieving CFI and TFI in excess of 0.95. Based on the collective evidence from our EFA and CFA, a three-factor model appears to provide "a parsimonious, substantively meaningful model that fits observed data adequately well" (MacCallum and Austin 2000). However, we note that there is no "true" model (Cudeck and Henly 1991; MacCallum and Austin 2000), and that construct validation is an ongoing process.

IV. RECOMMENDATIONS FOR SELECTING, ANALYZING, AND REPORTING DEPENDENT MEASURES

Our findings provide guidance on the selection, analysis, and reporting of dependent

[22] Factors 1 and 2 are even more highly correlated in the CFA. A likely reason for this is that many of the experimental materials used in the CFA focused on firms' financial performance and omitted any information that might cause Factor 2 to diverge from Factor 1.

measures for the experimental accounting literature on investor judgments. With a clearer understanding of the constructs that underlie common dependent measures, we recommend a five-step process that will help researchers select measures and analyze and report results.

Step 1: Rely on Theory to Identify the Conceptual Constructs of Interest

We recommend that researchers use a theory-driven approach to identify the conceptual constructs of interest prior to selecting investor-oriented dependent measures. Having performed an empirical validation of the dependent measures studied in prior work, our analyses point to three constructs that may often be of interest to researchers examining the judgments of nonprofessional investors.[23] There are two advantages to taking a theory-driven approach to construct identification and subsequent measure selection. First, clearly specifying conceptual constructs of interest helps place results into a broader conceptual framework, which should improve our ability to compare studies within the literature, link important findings to the archival literature, and generalize findings to real-world settings. Second, beginning with theory allows researchers to consider at the conceptual level how a given independent variable may affect different types of judgments made by investors.

Carefully applying Step 1 can help researchers select dependent measures that better operationalize their conceptual construct of interest, which should increase statistical power to detect hypothesized effects. For example, Factors 1 and 2 are positively correlated and should yield similar inferences (in expectation). However, when examining independent variables that are theorized to relate primarily to firm fundamental performance (e.g., earnings, cash flows, etc.), Factor 1 measures may yield greater statistical power than Factor 2 measures. In contrast,

[23] Certainly, future research is not limited to studying the three investor-oriented judgments we identify. Rather, our analysis highlights three constructs that have received significant attention in prior literature studying the judgments of nonprofessional investors.

when examining independent variables that are theorized to affect investor judgments, at least in part, through channels other than fundamental performance (e.g., implication of ESG, managerial communication style, etc.), measures that capture investors' holistic perceptions (Factor 2) may be more appropriate.

Fanning, Agoglia, and Piercey (2015) illustrate this recommendation well in their investigation of the effect of different disclosure thresholds for pending lawsuits on the judgments of nonprofessional investors. They state that because their "theory relates to how [their] manipulations of the litigation disclosures would influence investors' perceptions of disclosed litigation risk," they select a dependent measure that is closely related to capturing perceptions of litigation risk. They go on to say that other investment-related judgments may be impacted by their chosen independent variable, but "are likely to impound other noisy sources of variance because they are a less direct operational measure of [their] theoretical construct than litigation risk assessments." By following and explaining their theory-driven approach, they help readers understand their focus on investors' evaluations of risk and how they chose a dependent measure for detecting the effect predicted by their theory.

Step 2: Select a Set of Measures to Create a Measurement Scale

Our descriptive analysis reveals widespread variation in the number and types of dependent measures used in prior JDM research on investor judgment. For example, Figure 2 shows that 41 percent of studies only elicit one dependent measure when measuring investors' judgments. This can be problematic as single-item measures are subject to more measurement error because of random noise. As a result, researchers may allot time to projects that do not yield expected results. Conversely, researchers may obtain results that are not reliable when significant p-values occur by chance (Open Science Collaboration 2015; Harvey, Liu, and Zhu

2016; Wasserstein, Schirm, and Lazar 2019) or when the dependent measure and the independent variable share a common bias (Paulhus 1991). Alternatively, the other 59 percent of studies elicit between two and ten measures. This also can be problematic if the measures are selectively reported, inappropriately combined, and/or lead to inconsistent conclusions. Therefore, selecting the appropriate number of measures is of central importance.

Having identified a conceptual construct of interest, the number of associated measures that a researcher uses to capture it should depend on the degree to which the underlying construct is concrete and accessible versus abstract and inaccessible. Having reviewed the research on investor judgments, our impression is that the constructs of interest are often related to difficult evaluations based on complex information. As a result, "multiple items are likely to capture the variation in the construct with a degree of precision that a single item could not attain" (DeVellis 2016). [24] We recommend that in most cases, researchers should select multiple measures at an operational level that are intended to capture their construct of interest.

To help researchers decide on the specific dependent measures to select when studying one of the three constructs we identify, Table 5 presents the 34 measures from our analyses that loaded onto a single factor according to their EFA factor loadings.[25] We offer the table as an input into researchers' decision process and note that the organization does not reflect a ranking of measure appropriateness given a specific construct. Consistent with best practice, we urge researchers to carefully select dependent measures that capture different dimensions of the construct of interest. We caution against simply selecting dependent measures from Table 5

[24] Some constructs (e.g., age) are linked to relatively less ambiguous characteristics. As a result, well-designed single-item scales are often sufficient to accurately measure these constructs.
[25] Four measures cross-loaded on more than one factor in the EFA and were excluded from the CFA: DM11 (Rate the extent to which you agree with the following statement: "I feel very uncertain about investing in [Company] stock."); DM17 (How do you value [Company]'s stock?); DM18 (What do you believe the fundamental value for [Company] to be?); and DM19 (Indicate the value that you place on [Company]'s stock.).

based on their factor loadings alone.

Step 3: Provide Evidence of Unidimensionality

Construct validation is an ongoing process (Cronbach 1971). When using multiple measures, we recommend that researchers provide evidence of unidimensionality. In cases when researchers use at least four dependent measures and have an ex-ante prediction for the measurement model, researchers should perform and report the results of a CFA in lieu of an EFA. For example, researchers using the dependent measures from our analysis, but in new settings and across different participants, would ideally perform and report the results of a CFA using our results as the basis for their ex-ante measurement model prediction. Doing so helps to continue to validate the methods being implemented and increases confidence that the dependent measures used capture a single construct. This can also aid future researchers in determining whether a construct they wish to target can be captured via a set of measures used in prior work. We recommend that researchers perform a CFA when possible and report the model chi-square and associated p-value, the CFI, the TLI, the RMSEA and its 90 percent upper and lower interval bounds, and the SRMR. Each of these statistics provide a measure of model fit and are defined in Appendix A. Prescribed cutoffs for these statistics are provided in footnote 20 with more detailed information available in Hu and Bentler (1999).

In some cases, performing a CFA is not feasible or appropriate. For example, if researchers elicit three or fewer dependent measures to measure their dependent variable construct of interest, the resulting one-factor CFA model will either be just-identified or under-identified (meaning there are insufficient degrees of freedom). In these cases, model fit statistics are not meaningful. We find in our descriptive analysis that 88 percent of papers elicit three or fewer dependent measures, which likely explains why Cronbach’s alpha is the most cited reason for combining dependent measures in our sampled studies. Cronbach's alpha is a measure of scale *reliability* based on interitem *consistency* and can be useful to report. It is not, however, a sufficient indicator of scale *unidimensionality* (i.e., evidence that the selected dependent measures suitably represent one construct of interest).[26] This underscores the importance of our empirical analysis that uses a large set of dependent measures to identify different constructs of interest. When researchers create a measurement scale based on two or three dependent measures, we recommend that researchers rely on our results (and future validation evidence) when selecting dependent measures. Additionally, researchers should conduct an EFA on the dependent measures and report the percentage of variance explained by the first factor along with the Cronbach's alpha.

Step 4: Report Descriptive Statistics for Each Dependent Measure by Condition

In untabulated results from our descriptive analysis of the experimental investor judgment literature, we found considerable variation in how dependent measures are reported. Of the 90 articles we survey, 20 articles footnote at least one dependent measure. Factor 2 measures are footnoted most often (13 times), followed by Factor 1 measures (9 times) and Factor 3 measures (4 times). In some cases, the footnote is provided to inform readers that results are robust to using individual measures in place of a combined measure. However, two articles explicitly state that results are less significant with certain measures and eight articles footnote a measure as

[26] For example, the Cronbach's alpha for DM11, DM17, DM18, and DM19 is 0.87, even though these four measures capture aspects of all three factors. Similarly, the Cronbach's alpha for DM17, DM18, and DM19 is 0.93. While these three measures may appear on their face to capture a single construct, our analysis reveals that they cross-load onto Factors 1 and 2.

yielding insignificant results when analyzed individually.[27, 28] To increase the transparency and consistency with which dependent measures are reported, we recommend that researchers tabulate descriptive statistics (e.g., mean and standard deviation) by experimental condition for each of the primary dependent measures they elicit as part of their study.

Step 5: Combine Responses to Dependent Measures to Test Hypotheses

Our final recommendation is that researchers combine responses to the dependent measures that represent the same construct by taking the arithmetic mean. We recommend using the arithmetic mean because it is simple and minimizes both overfitting and researcher degrees of freedom in the choice of various combination approaches (e.g., regression scores, factor scores, principal components, etc.).[29] The combined metric should then serve as the primary dependent measure to test hypotheses. This approach should increase scale reliability and statistical power (DeVellis 2016).

Additional Considerations

We conclude this section by making a brief mention of two additional considerations researchers should keep in mind when eliciting responses from participants in experimental settings. First, researchers should be intentional in choosing the number of response options they provide when eliciting responses to dependent measures. There are two important factors to consider when determining the number of response options to provide: *variability* and

[27] The number of articles explicitly stating results are less significant is likely conservative. Many studies footnote individual dependent measures as providing "a similar pattern of results" or "similar inferences" without mentioning the statistical significance. We do not count these papers in the in-text tally but note it here as additional evidence regarding the variability with which dependent measure results are reported.
[28] The dependent measure that is footnoted most often as yielding less significant results is some variation of DM24: "How much of a $10,000 bonus would you invest in [Company]'s stock?"
[29] For example, as discussed by DiStefano, Zhu, and Mindrila (2009), "factor scores are sensitive to the factor extraction method and rotation method used…" (p. 5), and principal components should generally be used for data reduction purposes rather than being treated as latent variables.

discrimination (DeVellis 2016). The goal is to provide enough response options to capture meaningful variation in the judgments of participants, but not so many that participants cannot meaningfully discriminate between responses options. In our empirical analyses, we used all 7-point Likert-style response scales to minimize response bias attributable to methodological artifact.[30] While the choice of exactly seven response options is somewhat arbitrary, we note that participants' responses can be sensitive to the number of response options (and labels), which could in turn affect the results of an EFA or CFA.

Second, we urge researchers to be careful in their selection of measures used to provide process evidence. As discussed by Asay, Guggenmos, Kadous, Koonce, and Libby (2022), mediators and dependent measures should be both conceptually and operationally distinct. In our review, we identified 7 papers that report mediation analyses where the dependent measure and mediator load on the same construct in our analyses. In such cases, statistical mediation is all but assured, but the inferential value of the analyses is limited.[31]

V. DIRECTIONS FOR FUTURE RESEARCH

Having discriminated among the types of judgments made by nonprofessional investors and having provided recommendations based on best practices and our analyses, we now discuss areas where future research can continue to move the literature forward.

Investor Judgment and Decision-Making Framework

[30] Eutsler and Lang (2015) investigate the impact of the number of scale points and their labels for all measures presented to participants within accounting research in general. They find that a fully labeled 7-point scale is likely to provide the greatest benefit to researchers.

[31] To avoid this problem, we echo the recommendations of Asay et al. (2022) that researchers rely on established, well-validated measures of important theoretical constructs when providing process evidence. When no established measurement scale exists for a construct of interest, we recommend that researchers first seek to validate their measure before using it to provide process evidence.

Under traditional valuation approaches, firm value is a function of discounted expected future performance (e.g., Lee 1999; Lundholm and Sloan 2007), and investment preferences and decisions are a function of these valuations. However, in combination with our findings, the literature suggests a more complex and nuanced investor judgment and decision-making process. A well-developed descriptive model of the investor judgment and decision-making process could help synthesize prior work and improve our understanding of how investors evaluate investments and make decisions. Our findings may be helpful in developing such a framework.

For example, our empirical analysis suggests that nonprofessional investors discriminate among (1) expectations regarding a firm's future performance and value, (2) holistic perceptions of the firm (including nonpecuniary factors), and (3) evaluations of the risk associated with investing in a firm. Investors may also be interested in other factors not captured by the qualitative measures in our analyses (e.g., downstream investment decisions). A conceptual framework that articulates the interrelationship among the constructs we identify, the determinants and moderators of these constructs, and how these constructs lead to actual decisions by investors may be informative in identifying new areas of research.

Generalizability to Different Investor Samples

In our descriptive analysis, we attempt to be comprehensive by including studies that use participants that proxy for both nonprofessional (e.g., Amazon Mechanical Turk, undergraduate students, and graduate students) and professional investors (e.g., analysts). This aided our goal of generating an overinclusive set of dependent measures to inform our analyses. However, our factor analyses are limited to the judgments of nonprofessional investors as proxied by Amazon

Mechanical Turk workers.[32] Future research could examine the generalizability of the constructs we identify to alternative participant types. Distinguishing between nonprofessional and professional investors is likely to be particularly important given prior evidence that nonprofessional and professional participants engage in different processes when making judgments (Frederickson and Miller 2004).

Measurement Scale Development

In this paper, we provide recommendations for selecting, analyzing, and reporting dependent measures without constraining researchers to using a specific subset of dependent measures. However, scale development is common in psychology and other social sciences. Using our results as a foundation, future research could develop a standard subset of dependent measures that capture the different dimensions of the constructs we identify.

To illustrate, take the high correlation between our Factor 1 and Factor 2 as a starting point. The holistic perceptions of the firm captured through Factor 2 may also include some expectations regarding a firm's future performance and value. Separating the pecuniary and nonpecuniary components embedded in Factor 2 would aid researchers interested in studying either component on its own or in understanding the relationships between them. In fact, recent theoretical research in accounting and finance has begun to do exactly that by modeling investor behavior as a function of both *value* and *values* (e.g., Friedman and Heinle 2016; Pastor, Stambaugh, and Taylor 2021). Having standardized measurement scales for these related constructs would facilitate future experimental research in this area.

[32] We report that of the 90 articles in our survey, 12 used Amazon Mechanical Turk workers, 10 used undergraduate students, 59 used graduate students (including MBA students), and 25 used professional participants for at least one experiment. Several studies report that their results are robust to analyzing results by participant type separately or including participant type as a covariate (e.g., Kelly, Low, Tan, and Tan 2012; Harris, Hobson, and Jackson 2016; Elliott, Grant, and Rennekamp 2017; Emett and Nelson 2017).

Qualitative Versus Quantitative Measures

To draw conclusions about investors' judgments, researchers have used measures that are both quantitative (e.g., "provide an estimate of [X]") and qualitative (e.g., "what is the appropriate level {low to high} of [X]") in nature. In untabulated descriptive analysis, we find that researchers appear to be relatively more likely to select quantitative measures for more sophisticated participants, as a higher proportion of measures are quantitative when participants are graduate students or professionals than when participants are undergraduate students or recruited from the general public ($\chi^2(1) = 13.84$; $p < 0.001$). Studies with less sophisticated participants are also less likely to have at least one quantitative measure ($\chi^2(1) = 7.48$; $p = 0.006$). For practical purposes, our empirical analysis focuses solely on qualitative measures. However, future research can examine the extent to which qualitative and quantitative measures are able to provide similar inferences regarding a conceptual construct of interest.

VI. CONCLUSION

The literature on investor judgments and decisions has grown substantially in recent years. Our goal is to provide an overview of the dependent measures used in this literature, link those measures to potential constructs of interest, and provide researchers with a useful resource for making (and later supporting) their dependent measure choices. To do so, we survey the investor judgments literature, aggregate the most commonly used dependent measures, and use factor analyses to identify three underlying constructs related to the judgments of nonprofessional investors: (1) expectations regarding future firm performance and value, (2) holistic perceptions of the firm, and (3) evaluations of the risk associated with investing in the firm.

We also provide recommendations to help researchers select, analyze, and report dependent measures, as well as provide a number of important directions for future research that could increase our understanding of individual investor judgments, increase our confidence in the appropriateness of the dependent measures used to study these judgments, and provide additional insight to guide future choices of dependent measures. While our empirical analysis provides an important first step in assessing the validity of the dependent measures within the investor judgment literature, we believe future research can do more to assess the relationship between the experimental measures summarized here and the measures commonly used in archival work. Providing convergent and divergent validity between the measures used across these two methods would further assure the external validity of findings within the investor judgment experimental literature and help researchers generalize their results to real-world settings.

APPENDIX A. GLOSSARY OF TERMS

Below is a glossary of terms that may be useful to readers.

Common variance or communality	-	The fraction of the variance in each item (or dependent measure) that is accounted for by the latent constructs. The communality score is scaled such that a value of 1.0 indicates that the latent construct accounts for all the variation in the item. Uniqueness is the complement to communality.
Comparative Fit Index (CFI)	-	A measure of goodness of fit for a specified measurement model. The measurement model is compared to a baseline model in which there are no correlations between observed measures.
Confirmatory factor analysis	-	A special case of structural equation modeling that deals only with measurement models. It is a method of testing a priori hypotheses about the relationship between observed indicators (e.g., items) and theoretical latent constructs. CFA imposes explicit restrictions so that observed indicators relate with some (often just one) latent construct and not others.
Construct validity	-	The extent to which measurement scales map onto the latent constructs they are intended to represent.
Dimensionality	-	The number of latent constructs (or factors) needed to account for the correlation between items. In our case, three latent constructs are needed to account for the correlation between items intended to capture "investor judgments."
Eigenvalues	-	Indicates the proportion of the total variance among several correlated items that is accounted for by a more basic, underlying latent construct.
Exploratory factor analysis	-	A method for discovering a small set of underlying latent constructs from a large set of observable indicators (e.g., items).
Face (translational) validity	-	The extent to which items appear to be appropriate for capturing a latent construct based on their content, regardless of whether they actually are from an empirical standpoint (Bloomfield, Nelson, and Soltes 2016; DeVellis 2016).
Factor loadings	-	The correlation between an item and a latent construct (or factor).
Items	-	Individual questions that elicit responses from participants resulting in observable indictors of a latent construct. Each of the 38 dependent measures we study are an item.
Item complexity	-	Indicates the degree to which an item (or dependent measure) captures a single construct. The item complexity score will equal 1.0 if an item loads on exactly one factor, 2.0 if it evenly loads on two factors, and so on.
Latent construct	-	A theoretical variable of interest that is not directly observable

		but rather approximated through various observable indicators (e.g., items). Sometimes referred to as a latent factor.
Measurement model	-	A model that specifies the relationship between observed indicators (e.g., items) and theoretical latent constructs.
Measurement scales	-	A collection of items that capture a latent construct (or factor). There is no single correct measurement scale. However, a measurement scale should be validated to provide greater confidence that the intended construct is being appropriately measured.
Model chi-square	-	A measure of goodness of fit for a specified measurement model. Tests the null hypothesis that the model is a perfect fit for the data.
Root Mean Square Error of Approximation (RMSEA)	-	A measure of goodness of fit for a specified measurement model. Evaluates the measurement model as being an approximate fit for the data, rather than a perfect fit.
Standardized Root Mean Square Residual (SRMR)	-	A measure of goodness of fit for a specified measurement model. An index of the average standardized residuals between the observed and predicted covariance matrices.
Structural validity	-	The degree to which responses to the measurement instrument are an adequate reflection of the dimensionality of the construct being measured.
Tucker-Lewis Index (TLI)	-	A measure of goodness of fit for a specified measurement model. Like CFI, the measurement model is compared to a baseline model in which there are no correlations between observed measures. However, TLI penalizes overly complex model specifications.
Unique variance or uniqueness	-	The fraction of the variance in each item (or dependent measure) that is unaccounted for by the latent constructs. Communality is the complement to uniqueness.

APPENDIX B. EXPERIMENTAL ACCOUNTING ARTICLES ON INVESTOR JUDGMENT

Below is the list of 90 financial experimental research articles we identified as investor judgment studies. The articles are listed chronologically first, then alphabetically. Citations in **bold** typeface are the ones with experimental materials that we used for our confirmatory factor analysis. Articles with an asterisk were excluded from our confirmatory factor analysis because they provided experimental materials that were longer than a single page.

Hirst, Koonce, and Simko (1995)

Hopkins (1996)

Maines, McDaniel, and Harris (1997)

Hirst and Hopkins (1998)

Kennedy, Mitchell, and Sefcik (1998)

Lipe (1998)

Hirst, Koonce, and Miller (1999)

Libby and Tan (1999)

Tuttle and Burton (1999)

Hopkins, Houston, and Peters (2000)

Maines and McDaniel (2000)*

Hodge (2001)

Sedor (2002)

Hirst, Jackson, and Koonce (2003)

Fredrickson and Miller (2004)

Hirst, Hopkins, and Whalen (2004)

Hodge, Kennedy, and Maines (2004)

Barton and Mercer (2005)

Koonce, Lipe, and McAnally (2005)

Koonce, McAnally, and Mercer (2005)

Krische (2005)

Mercer (2005)

Elliott (2006)

Frederickson, Hodge, and Pratt (2006)

Kadous, Krische, and Sedor (2006)

Miller (2006)

Hales (2007)

Han and Tan (2007)

Hirst, Koonce, Venkataraman (2007)

Hodder, Hopkins, and Wood (2008)

Koonce, Lipe, and McAnally (2008)

Pinello (2008)

Clor-Proell (2009)

Hewitt (2009)

Tan and Tan (2009)

Elliott, Krische, and Peecher (2010)

Han and Tan (2010)

Hodge, Hopkins, and Wood (2010)

Koonce and Lipe (2010)

Koonce, Williamson, and Winchel (2010)

Rose, Norman, and Rose (2010)

Elliott, Hobson, and Jackson (2011)

Hales, Kuang, and Venkataraman (2011)

Koonce, Nelson, and Shakespeare (2011)

Tan and Koonce (2011)

Elliott, Hodge, and Sedor (2012)

Kadous, Koonce, and Thayer (2012)

Kelly, Low, Tan, and Tan (2012)

Maletta and Zhang (2012)

Rennekamp (2012)

Chen and Tan (2013)

Bonner, Clor-Proell, and Koonce (2014)

Clor-Proell, Proell, and Warfield (2014)*

Elliott, Jackson, Peecher, and White (2014)*

Tan, Wang, and Zhou (2014)*

Anderson, Brown, Hodder, and Hopkins (2015)

Elliott, Rennekamp, and White (2015)

Fanning, Agoglia, and Piercey (2015)

Hewitt, Tarca, and Yohn (2015)

Koonce, Miller, and Winchel (2015)

Lachmann, Stefani, and Wohrmann (2015)

Nelson and Rupar (2015)

Tan, Wang, and Zhou (2015)

Winchel (2015)

Chen, Han, and Tan (2016)

Harris, Hobson, and Jackson (2016)

Asay, Elliott, and Rennekamp (2017)

Dong, Lui, and Wong-on-Wing (2017)

Elliott, Grant, and Rennekamp (2017)

Emett and Nelson (2017)*

Erickson, Hewitt, and Maines (2017)

Kelly and Tan (2017)

Koonce and Lipe (2017)

Rupar (2017)

Asay and Hales (2018)

Asay, Libby, and Rennekamp (2018)

Cade (2018)

Elliott, Grant, and Hodge (2018)

Grant, Hodge, Sinha (2018)

Kelton and Montague (2018)

Tan and Yu (2018)

Tang and Venkataraman (2018)

Cardinaels, Hollander, and White (2019)

Chen and Loftus (2019)

Emett (2019)

He, Tan, Yeo, and Zhang (2019)

Koonce, Leitter, and White (2019)

Tan, Wang, and Yoo (2019)

Elliott, Fanning, and Peecher (2020)

Guiral, Moon, Tan, and Yu (2020)

APPENDIX C. EXPERIMENTAL VIGNETTES FOR EXPLORATORY FACTOR ANALYSIS

Below is the information provided to experimental participants prior to eliciting responses to the various dependent measures. The first bullet point is our expected future earnings manipulation, the second bullet point is our investment risk manipulation, and the third bullet point is our firm values manipulation. The 'High' level of the manipulation is as read. The 'Low' level of the manipulation is included in parentheses. The order of the bullet points and assignment to one of the eight vignettes is randomized.

Background Information:

For this study, you will assume the role of an investor making judgments about a company. You will read a brief summary about a company and then respond to a series of questions. Your responses should be based on the information provided and your own preferences and experiences.

Company Information:

Today, you will be making judgments about Kappa Corp. Kappa Corp is a publicly traded company. Please review the information below before proceeding to the next page:

- Kappa's managers forecast earnings growth to be 15% (1%) next year. The average of individual analyst forecasts also projects strong (weak) earnings growth for Kappa over the next several years.

- Growth in demand for Kappa's products is relatively volatile (stable), causing future earnings to be somewhat unpredictable (fairly predictable). As a result, individual analyst forecasts of earnings growth are considerably different (generally similar), with some significantly (slightly) lower and some significantly (slightly) higher than the average forecast of earnings growth.

- Kappa is widely viewed as having a positive (negative) impact on the environment and is known for treating its employees well (poorly). Further, recent news coverage has praised (criticized) Kappa for supporting (damaging) small communities and opposing (paying) bribes to foreign governments.

Before proceeding to the next page, please check the box below to indicate that you have carefully read the information about Kappa Corp.

☐ I have read the information about Kappa Corp. carefully

References

Anderson, S. B., J. L. Brown, L. E. Hodder, and P. Hopkins. 2015. The effect of alternative accounting measurement bases on investors' assessments of managers' stewardship. *Accounting, Organizations and Society* 46: 100–114.

Asay, H. S., W. B. Elliott, and K. M. Rennekamp. 2017. Disclosure readability and the sensitivity of investors' valuation judgments to outside information. *The Accounting Review* 92 (4): 1–25.

Asay, H. S., R. D. Guggenmos, K. Kadous, L. Koonce, and R. Libby. 2021. Theory testing and process evidence in accounting experiments. *The Accounting Review* (forthcoming).

Asay, H. S., and J. Hales. 2018. Disclaiming the future: Investigating the impact of cautionary disclaimers on investor judgments before and after experiencing economic loss. *The Accounting Review* 93 (4): 81–99.

Asay, H. S., R. Libby, and K. M. Rennekamp. 2018. Do features that associate managers with a message magnify investors' reactions to narrative disclosures? *Accounting, Organizations and Society* 68-69: 1–14.

Barton, J., and M. Mercer. 2005. To blame or not to blame: Analysts' reactions to external explanations for poor financial performance. *Journal of Accounting and Economics* 39 (3): 509–533.

Bentley, J. W. 2021. Improving the statistical power and reliability of research using Amazon Mechanical Turk. *Accounting Horizons* 35 (4): 45–62.

Bloomfield, R., M. W. Nelson, and E. Soltes. 2016. Gathering data for archival, field, survey, and experimental accounting research. *Journal of Accounting Research* 54 (2): 341–395.

Bonner, S. E., S. M. Clor-Proell, and L. Koonce. 2014. Mental accounting and disaggregation based on the sign and relative magnitude of income statement items. *The Accounting Review* 89 (6): 2087–2114.

Cade, N. L. 2018. Corporate social media: How two-way disclosure channels influence investors. *Accounting, Organizations and Society* 68-69: 63–79.

Cardinaels, E., S. Hollander, and B. J. White. 2019. Automatic summarization of earnings releases: attributes and effects on investors' judgments. *Review of Accounting Studies* 24 (3): 860–890.

Cattell, R. B. 1966. The Scree Test for the Number of Factors. *Multivariate Behavioral Research* 1 (2).

Cattell, R. B. 1978. *The Scientific Use of Factor Analysis*. New York, NY: Plenum.

Center for Audit Quality. 2019. Mainstreet Investor Survey. Retrieved (2022) from (https://www.thecaq.org/2019-main-street-investor-survey/).

Chen, W., and H. T. Tan. 2013. Judgment effects of familiarity with an analyst's name. *Accounting, Organizations and Society* 38 (3): 214–227.

Chen, W., J. Han, and H. T. Tan. 2016. Investor reactions to management earnings guidance attributions: The effects of news valence, attribution locus, and outcome controllability. *Accounting, Organizations and Society* 55: 83–95.

Chen, Z. and S. Loftus. 2019. Multi-method evidence on investors' reactions to managers' self-inclusive language. *Accounting, Organizations and Society* 79: 1–19.

Clor-Proell, S. M. 2009. The effects of expected and actual accounting choices on judgments and decisions. *The Accounting Review* 84 (5): 1465–1493.

Clor-Proell, S. M., R. D. Guggenmos, and K. Rennekamp. 2020. Mobile devices and investment news apps: The effects of information release, push notification, and the fear of missing out. *The Accounting Review* 95 (5): 95–115.

Clor-Proell, S. M., C. A. Proell, and T. D. Warfield. 2014. The effects of presentation salience and measurement subjectivity on nonprofessional investors' fair value judgments. *Contemporary Accounting Research* 31 (1): 45–66.

Comrey, A. L., and H. B. Lee. 1992. Interpretation and application of factor analytic results. In *A first course in factor analysis*, edited by A. L. Comrey, and H.B. Lee. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, T. D., and D. T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues in Field Settings*. Houghton Mifflin.

Cronbach, L. J. 1971. Test validation. In *Educational measurement*, edited by R. L. Thorndike. Washington, D.C.: American Council on Education.

Cudeck, R., and S. J. Henly. 1991. Model selection in covariance structures analysis and the "problem" of sample size: A clarification. *Psychological Bulletin* 109 (3): 512.

DeVellis, R. F. 2016. *Scale Development Theory and Applications*. Fourth edition. SAGE Publication 4.

DiStefano, C., M. Zhu, and D. Mindrila. 2009. Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation* 14 (1): 20.

Dong, L., G. Lui, and B. Wong-On-Wing. 2017. Unintended consequences of forecast disaggregation: A multi-period perspective. *Contemporary Accounting Research* 34 (3): 1580–1595.

Elliott, W. B. 2006. Are investors influenced by pro forma emphasis and reconciliations in earnings announcements? *The Accounting Review* 81 (1): 113–133.

Elliott, W. B., K. Fanning, and M. E. Peecher. 2020. Do investors value higher financial reporting quality, and can expanded audit reports unlock this value? *The Accounting Review* 95 (2): 141–165.

Elliott, W. B., S. M. Grant, and F. D. Hodge. 2018. Negative news and investor trust: The role of $Firm and #CEO twitter use. *Journal of Accounting Research* 56 (5): 1483–1519.

Elliott, W. B., S. M. Grant, and K. M. Rennekamp. 2017. How disclosure features of corporate social responsibility reports interact with investor numeracy to influence investor judgments. *Contemporary Accounting Research* 34 (3): 1596–1621.

Elliott, W. B., J. L. Hobson, and K. E. Jackson. 2011. Disaggregating management forecasts to reduce investors' susceptibility to earnings fixation. *The Accounting Review* 86 (1): 185–208.

Elliott, W. B., F. D. Hodge, J. J. Kennedy, and M. Pronk. 2007. Are MBA students a good proxy for nonprofessional investors? *The Accounting Review* 82 (1): 139–168.

Elliott, W. B., F. D. Hodge, and L. Sedor. 2012. Using online video to announce a restatement: Influences on investment decisions and the mediating role of trust. *The Accounting Review* 87 (2): 513–535.

Elliott, W. B., K. E. Jackson, M. E. Peecher, and B. J. White. 2014. The unintended effect of corporate social responsibility performance on investors' estimates of fundamental value. *The Accounting Review* 89 (1): 275–302.

Elliott, W. B., S. D. Krische, and M. E. Peecher. 2010. Expected mispricing: The joint influence of accounting transparency and investor base. *Journal of Accounting Research* 48 (2): 343–381.

Elliott, W. B., K. M. Rennekamp, and B. J. White. 2015. Does concrete language in disclosures increase willingness to invest? *Review of Accounting Studies* 20: 839–865.

Emett, S.A. 2019. Investor reaction to disclosure of past performance and future plans. *The Accounting Review* 94 (5): 165–188.

Emett, S. A., and M. W. Nelson. 2017. Reporting accounting changes and their multi-period effects. *Accounting, Organizations and Society* 57: 52–72.

Erickson, D. A., M. A. Hewitt, and L. A. Maines. 2017. Do investors perceive low risk when earnings are smooth relative to the volatility of operating cash flows? *The Accounting Review* 92 (3): 137–154.

Eutsler, J. and B. Lang. 2015. Rating scales in accounting research: The impact of scale points and labels. *Behavioral Research in Accounting*, Vol 27(2): 35–51.

Everitt, B. S. 1975. Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*. 126 (3): 237–240.

Fama, E. F. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49 (3).

Fanning, K., C. P. Agoglia, and M. D. Piercey. 2015. Unintended consequences of lowering disclosure thresholds. *The Accounting Review* 90 (1): 301–320.

Farrell, A. M., J. H. Grenier, and J. Leiby. 2017. Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review* 92 (1).

Flake, J. K., J. Pek, and E. Hehman. 2017. Construct validation in social and personality research: current practice and recommendations. *Social Psychological and Personality Science* 8 (4): 370–378.

Frederickson, J. R., F. D. Hodge, and J. H. Pratt. 2006. The evolution of stock option accounting: Disclosure, voluntary recognition, mandated recognition, and management disavowals. *The Accounting Review* 81 (5): 1073–1093.

Frederickson, J. R., and J. S. Miller. 2004. The effects of pro forma earnings disclosures on analysts' and nonprofessional investors' equity valuation judgments. *The Accounting Review* 79 (3): 667–686.

Friedman, H. L., and M. S. Heinle. 2016. Taste, information, and asset prices: Implications for the valuation of CSR. *Review of Accounting Studies* 21 (3): 740–767.

Gorsuch, R. L. 1983. *Factor analysis*. Second edition. Hillsdale, NJ: Erlbaum Associates.

Grant, S. M., F. D. Hodge, and R. K. Sinha. 2018. How disclosure medium affects investor reactions to CEO bragging, modesty, and humblebragging. *Accounting, Organizations and Society* 68–69: 118–134.

Guilford, J. P. 1954. *Psychometric methods*. Second edition. New York, NY: McGraw-Hill.

Guiral, A., D. Moon, H. T. Tan, and Y. Yu. 2020. What drives investor response to CSR performance reports? *Contemporary Accounting Research*.

Hales, J. 2007. Directional preferences, information processing, and investors' forecasts of earnings. *Journal of Accounting Research* 45 (3): 607–628.

Hales, J., X. Kuang, and S. Venkataraman. 2011. Who believes the hype? An experimental examination of how language affects investor judgments. *Journal of Accounting Research* 49 (1): 223–255.

Han, J., and H. T. Tan. 2007. Investors' reactions to management guidance forms: The influence of multiple benchmarks. *The Accounting Review* 82 (2): 521–543.

Han, J., and H. T. Tan. 2010. Investors' reactions to management earnings guidance: The joint effect of investment position, news valence, and guidance form. *Journal of Accounting Research* 48 (1): 81–104.

Harris, L. L., J. L. Hobson, and K. E. Jackson. 2016. The effect of investor status on investors' susceptibility to earnings fixation. *Contemporary Accounting Research* 33 (1): 152–171.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. … and the cross-section of expected returns. *The Review of Financial Studies* 29 (1): 5–68.

He, Y., H. T. Tan, F. Yeo, and J. Zhang. 2019. When do qualitative risk disclosures backfire? The effects of a mismatch in hedge disclosure formats on investors' judgments. *Contemporary Accounting Research* 36 (4): 2093–2112.

Hewitt, M. 2009. Improving investors' forecast accuracy when operating cash flows and accruals are differentially persistent. *The Accounting Review* 84 (6): 1913–1931.

Hewitt, M. L., A. L. Tarca, and T. L. Yohn. 2015. The effect of measurement subjectivity classifications on analysts' use of persistence classifications when forecasting earnings items. *Contemporary Accounting Research* 32 (3): 1000–1023.

Hirst, D. E., and P. E. Hopkins. 1998. Comprehensive income reporting and analysts' valuation judgments. *Journal of Accounting Research* 36: 47–75.

Hirst, D. E., P. E. Hopkins, and J. M. Wahlen. 2004. Fair values, income measurement, and bank analysts' risk and valuation judgments. *The Accounting Review* 79 (2): 453–472.

Hirst, D. E., K. E. Jackson, and L. R. Koonce. 2003. Improving financial reports by revealing the accuracy of prior estimates. *Contemporary Accounting Research* 20 (1): 165–199.

Hirst, D. E., L. E. Koonce, and J. E. Miller. 1999. The joint effect of management's prior forecast accuracy and the form of its financial forecasts on investor judgment. *Journal of Accounting Research* 37: 101–124.

Hirst, D. E., L. J. Koonce, and P. J. Simko. 1995. Investor reactions to financial analysts' research reports. *Journal of Accounting Research* 33 (2): 335–351.

Hirst, D. E., L. E. Koonce, and S. E. Venkataraman. 2007. How disaggregation enhances the credibility of management earnings forecasts. *Journal of Accounting Research* 45 (4): 811–837.

Hodder, L. E., P. A. Hopkins, and D. Wood. 2008. The effects of financial statement and informational complexity on analysts' cash flow forecasts. *The Accounting Review* 83 (4): 915–

956.

Hodge, F. D. 2001. Hyperlinking unaudited information to audited financial statements: effects on investor judgments. *The Accounting Review* 76 (4): 675–691.

Hodge, F. D., P. E. Hopkins, and D. A. Wood. 2010. The effects of financial statement information proximity and feedback on cash flow forecasts. *Contemporary Accounting Research* 27 (1): 101–133.

Hodge, F. D., J. J. Kennedy, and L. A. Maines. 2004. Does search-facilitating technology improve the transparency of financial reporting? *The Accounting Review* 79 (3): 687–703.

Hopkins, P. E. 1996. The effect of financial statement classification of hybrid financial instruments on financial analysts' stock price judgments. *Journal of Accounting Research* 34: 33–50.

Hopkins, P. E., R. W. Houston, and M. F. Peters. 2000. Purchase, pooling, and equity analysts' valuation judgments. *The Accounting Review* 75 (3): 257–281.

Horn, J. L. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30 (2).

Hu, L. T., and P. M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6 (1).

Johnson, J.A., J. Theis, A. Vitalis, and D. Young. 2020. The influence of firms' emissions management strategy disclosures on investors' valuation judgments. *Contemporary Accounting Research* 37 (2): 642–664.

Kadous, K., L. Koonce, and J. M. Thayer. 2012. Do financial statement users judge relevance based on properties of reliability? *The Accounting Review* 87 (4): 1335–1356.

Kadous, K., S. D. Krische, and L. M. Sedor. 2006. Using counter-explanation to limit analysts' forecast optimism. *The Accounting Review* 81 (2): 377–397.

Kahneman, D., and S. Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman. Cambridge University Press.

Kaiser, H. F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20 (1).

Kelly, K., B. Low, H. T. Tan, and S. K. Tan. 2012. Investors' reliance on analysts' stock recommendations and mitigating mechanisms for potential overreliance. *Contemporary Accounting Research* 29 (3): 991–1012.

Kelly, K., and H. T. Tan. 2017. Mandatory management disclosure and mandatory independent audit of internal controls: Evidence of configural information processing by investors. *Accounting, Organizations and Society* 56: 1–20.

Kelton, A. S., and N. R. Montague. 2018. The unintended consequences of uncertainty disclosures made by auditors and managers on nonprofessional investor judgments. *Accounting, Organizations and Society* 65: 44–55.

Kennedy, J., T. Mitchell, and S. E. Sefcik. 1998. Disclosure of contingent environmental liabilities: Some unintended consequences? *Journal of Accounting Research* 36 (2): 257–277.

Kline, P. 1979. *Psychometrics and psychology*. London, U.K.: Acaderric Press.

Koonce, L., Z. Leitter, and B. J. White. 2019. Linked balance sheet presentation. *Journal of Accounting and Economics* 68 (1): 1–16.

Koonce, L., and M. G. Lipe. 2010. Earnings trend and performance relative to benchmarks: How consistency influences their joint use. *Journal of Accounting Research* 48 (4): 859–884.

Koonce, L., and M. G. Lipe. 2017. Firms with inconsistently signed earnings surprises: Do potential investors use a counting heuristic? *Contemporary Accounting Research* 34 (1): 292–313.

Koonce, L., M. G. Lipe, and M. L. McAnally. 2005. Judging the risk of financial instruments: Problems and potential remedies. *The Accounting Review* 80 (3): 871–895.

Koonce, L., M. G. Lipe, and M. L. McAnally. 2008. Investor reactions to derivative use and outcomes. *Review of Accounting Studies* 13 (4): 571–597.

Koonce, L., M. L. McAnally, and M. Mercer. 2005. How do investors judge the risk of financial items? *The Accounting Review* 80 (1): 221–241.

Koonce, L., J. Miller, and J. Winchel. 2015. The effects of norms on investor reactions to derivative use. *Contemporary Accounting Research* 32 (4): 1529–1554.

Koonce, L., K. Nelson, and C. M. Shakespeare. 2011. Judging the relevance of fair value for financial instruments. *The Accounting Review* 86 (6): 2075–2098.

Koonce, L., M. Williamson, and J. Winchel. 2010. Consensus information and nonprofessional investors' reaction to the revelation of estimate inaccuracies. *The Accounting Review* 85 (3): 979–1000.

Kothari, S. P. 2001. Capital markets research in accounting. *Journal of Accounting and Economics* 31 (1–3).

Krische, S. 2005. Investors' evaluations of strategic prior-period benchmark disclosures in earnings announcements. *The Accounting Review* 80 (1): 243–268.

Krische, S. 2019. Investment experience, financial literacy, and investment-related judgments. *Contemporary Accounting Research* 36 (3): 1634–1668.

Lachmann, M., U. Stefani, and A. Wöhrmann. 2015. Fair value accounting for liabilities: Presentation format of credit risk changes and individual information processing. *Accounting, Organizations and Society* 41: 21–38.

Lee, CM. 1999. Accounting-based valuation: Impact on business practices and research. *Accounting Horizons* 13 (4): 413–425.

Libby, R., R. Bloomflield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 775–810.

Libby, R. and S. A. Emett. 2014. Earnings presentation effects on manager reporting choices and investor decisions, *Accounting and Business Research* 44(4): 410–438.

Libby, R., and H. T. Tan. 1999. Analysts' reactions to warnings of negative earnings surprises. *Journal of Accounting Research* 37 (2): 415–435.

Lipe, M. G. 1998. Individual investors' risk judgments and investment decisions: The impact of accounting and market data. *Accounting, Organizations and Society* 23 (7): 625–640.

Loevinger, J. 1957. Objective tests as instruments of psychological theory. *Psychological Reports* 3 (3).

Lundholm, R., and R. Sloan. 2007. *Equity Valuation and Analysis with eVal*. Second edition. McGraw-Hill/Irwin.

MacCallum, R.C., and J.T. Austin. 2000. Applications of structural equation modeling in psychological research. *Annual Review of Psychology* 51.

Maines, L. A., and L. S. McDaniel. 2000. Effects of comprehensive-income characteristics on nonprofessional investors' judgments: The role of financial-statement presentation format. *The Accounting Review* 75 (2): 179–207.

Maines, L. A., L. S. McDaniel, and M. S. Harris. 1997. Implications of proposed segment reporting standards for financial analysts' investment judgments. *Journal of Accounting Research* 35: 1–24.

Maletta, M., and Y. Zhang. 2012. Investor reactions to contrasts between the earnings preannouncements of peer firms. *Contemporary Accounting Research* 29 (2): 361–381.

Mercer, M. 2004. How do investors assess the credibility of management disclosures? *Accounting Horizons* 18 (3).

Mercer, M. 2005. The fleeting effects of disclosure forthcomingness on management's reporting credibility. *The Accounting Review* 80 (2): 723–744.

Miller, J. 2006. Unintended effects of preannouncements on investor reactions to earnings news. *Contemporary Accounting Research* 23 (4): 1073–1103.

Nelson, M. W., and K. Rupar. 2015. Numerical formats within risk disclosures and the moderating effect of investors' concerns about management discretion. *The Accounting Review* 90 (3): 1149–1168.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349 (6251): aac4716.

Pástor, Ľ., R. F. Stambaugh, and L. A. Taylor. 2021. Sustainable investing in equilibrium. *Journal of Financial Economics* 142 (2): 550–571.

Paulhus, D. L. 1991. Measurement and control of response bias. In *Measures of personality and social psychological attitudes*, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman. Academic Press.

Pinello, A. S. 2008. Investors' differential reaction to positive versus negative earnings surprises. *Contemporary Accounting Research* 25 (3): 891–920.

Rennekamp, K. 2012. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research* 50 (5): 1319–1354.

Rose, J., C. Norman, and A. Rose. 2010. Perceptions of investment risk associated with material control weakness pervasiveness and disclosure detail. *The Accounting Review* 85 (5): 1787–1807.

Rupar, K. 2017. Significance of forecast precision: The importance of investors' expectations. *Contemporary Accounting Research* 34 (2): 849–870.

Sedor, L. M. 2002. An explanation for unintentional optimism in analysts' earnings forecasts. *The Accounting Review* 77 (4): 731–753.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Summers, S. L., and D. A. Wood. (2018). Accounting Research Ranking. Retrieved (2018) from (http://www.byuaccounting.net/rankings/univrank/rank_university.php?sortorder=ranking6andqurank=All.)

Tan, H. T., and S. K. Tan. 2009. Investors' reactions to management disclosure corrections: Does presentation format matter? *Contemporary Accounting Research* 26 (2): 605–626.

Tan, H. T., E. Y. Wang, and G. S. Yoo. 2019. Who likes jargon? The joint effect of jargon type and industry knowledge on investors' judgments. *Journal of Accounting and Economics* 67 (2-3): 416–437.

Tan, H. T., E. Wang, and B. Zhou. 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research* 52 (1): 273–302.

Tan, H. T., E. Wang, and B. Y. Zhou. 2015. How does readability influence investors' judgments? Consistency of benchmark performance matters. *The Accounting Review* 90 (1): 371–393.

Tan, H. T., and Y. Yu. 2018. Management's responsibility acceptance, locus of breach, and investors' reactions to internal control reports. *The Accounting Review* 93 (6): 331–355.

Tan, S. K., and L. Koonce. 2011. Investors' reactions to retractions and corrections of management earnings forecasts. *Accounting, Organizations and Society* 36 (6): 382–397.

Tang, M., and S. Venkataraman. 2018. How patterns of past guidance provision affect investor judgments: The joint effect of guidance frequency and guidance pattern consistency. *The Accounting Review* 93 (3): 327–348.

Thaler, R. 1999. The end of behavioral finance: Why behavioral finance cannot be dismissed. *Financial Analysts Journal* 55 (6).

Tuttle, B., and F. G. Burton. 1999. The effects of a modest incentive on information overload in an investment analysis task. *Accounting, Organizations and Society* 24 (8): 673–687.

Velicer, W. F., C. A. Eaton, and J. L. Fava. 2000. Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In *Problems and Solutions in Human Assessment*.

Wasserstein, R. L., A. L. Schirm, and N. A. Lazar. 2019. Moving to a world beyond "p< 0.05". *The American Statistician* 73 (sup1): 1–19.

Winchel, J. 2015. Investor reactions to the ambiguity and mix of positive and negative argumentation in favorable analyst reports. *Contemporary Accounting Research* 32 (3): 973–999.

Zwick, W. R., and W. F. Velicer. 1986. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin* 99 (3).

FIGURE 1: FREQUENCY OF JUDGMENT TYPE MEASUREMENT

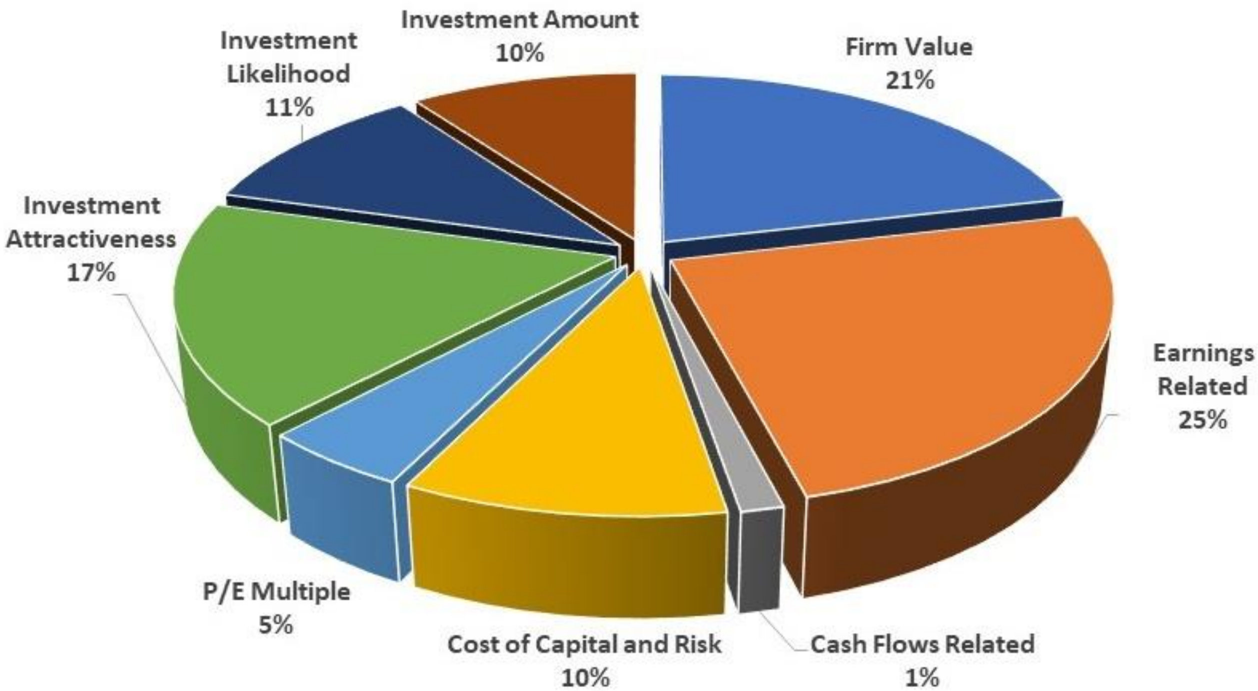

Figure 1 depicts the frequency with which different types of investor judgments were measured within our sample of research papers. For papers with multiple experiments, these counts include only the dependent measures for the first experiment if the *same* primary dependent measures are elicited in subsequent experiments. However, these counts include the dependent measures for subsequent experiments that elicit *different* primary dependent measures.

FIGURE 2: NUMBER OF PRIMARY DEPENDENT MEASURES ELICITED

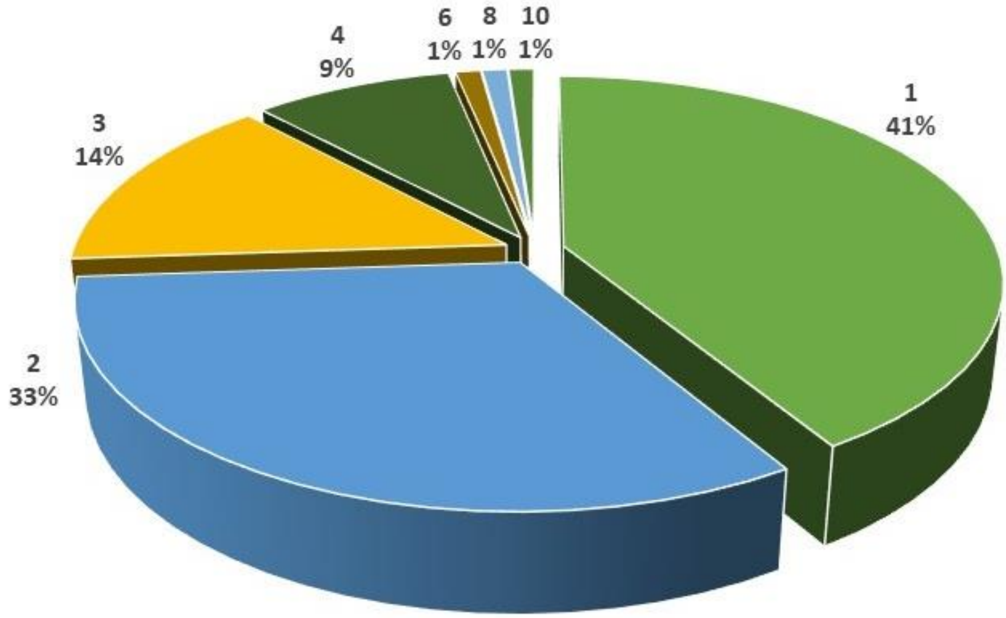

Figure 2 depicts the number of primary dependent measures elicited in each research paper in our sample. For papers with multiple experiments, these counts include only the dependent measures for the first experiment if the *same* primary dependent measures are elicited in subsequent experiments. However, these counts include the dependent measures for subsequent experiments that elicit *different* primary dependent measures.

TABLE 1. DEPENDENT MEASURES USED IN EXPLORATORY FACTOR ANALYSIS

Dependent Measures Reference Sheet		**Response Scale**		
Ref.	**Measure**	**1**	**Mid**	**7**
DM1	[Company]'s future earnings performance will be _______.	Very Weak	About Average	Very Strong
DM2	Rate the extent to which you agree with the following statement: "[Company]'s earnings performance will be strong in the near future."	Strongly Disagree	Neither Disagree Nor Agree	Strongly Agree
DM3	[Company]'s earnings potential over the next year is _______.	Very Weak	About Average	Very Strong
DM4	How much do you think [Company]'s earnings will grow in the future?	Far Below Average	About Average	Far Above Average
DM5	You forecast [Company]'s earnings for the next year to be _______.	Very Weak	About Average	Very Strong
DM6	You forecast [Company]'s earnings growth rate for the next year to be _______.	Very Weak	About Average	Very Strong
DM7	You forecast [Company]'s cash flows for the next year to be _______.	Very Weak	About Average	Very Strong
DM8	How risky do you consider an investment in [Company] stock to be?	Very Low Risk	About Average Risk	Very High Risk
DM9	What is the risk of [Company] experiencing a moderate stock price decline within the next year?	Very Low Risk	About Average Risk	Very High Risk
DM10	Rate the risk associated with investing in [Company] stock.	Very Low Risk	About Average Risk	Very High Risk
DM11	Rate the extent to which you agree with the following statement: "I feel very uncertain about investing in [Company] stock."	Strongly Disagree	Neither Disagree Nor Agree	Strongly Agree
DM12	The investment risk of [Company] in the context of a diversified portfolio is _______.	Very Low Risk	About Average Risk	Very High Risk
DM13	Rate the extent to which you agree with the following statement: "Over the next 12 months, the stock price could increase significantly."	Strongly Disagree	Neither Disagree Nor Agree	Strongly Agree
DM14	Rate the extent to which you agree with the following statement: "Over the next 12 months, the stock price could decrease significantly."	Strongly Disagree	Neither Disagree Nor Agree	Strongly Agree
DM15	What do you believe is an appropriate common stock valuation for [Company]?	Very Low Value	About Average Value	Very High Value
DM16	What do you expect [Company]'s stock price per share to be next year?	Very Low	About Average	Very High

TABLE 1 (CONTINUED)

Ref.	Measure	1	Mid	7
DM17	How do you value [Company]'s stock?	Very Low Value	About Average Value	Very High Value
DM18	What do you believe the fundamental value for [Company] to be?	Very Low Value	About Average Value	Very High Value
DM19	Indicate the value that you place on [Company]'s stock.	Very Low Value	About Average Value	Very High Value
DM20	I believe that other stock market participants would value [Company]'s stock at a _______.	Very Low Value	About Average Value	Very High Value
DM21	What common stock valuation do you think potential investors would place on [Company] stock?	Very Low Value	About Average Value	Very High Value
DM22	How do you believe investors would perceive the value of [Company] stock?	Very Low Value	About Average Value	Very High Value
DM23	Provide a judgment about the appropriate valuation of [Company] stock.	Very Low Value	About Average Value	Very High Value
DM24	How much of a $10,000 bonus would you invest in [Company]'s stock?	None	Half	All
DM25	Provide a buy/sell/hold recommendation for [Company]'s stock.	Strong Sell	Hold	Strong Buy
DM26	What would be your preferred investment position in [Company]'s stock?	Large Short Position	No Position	Large Long Position
DM27	Assume you were placing a $10,000 bet on [Company]'s future stock price. Would you bet on [Company]'s stock price increasing or decreasing over the next year?	Definitely Decreasing	Not Sure	Definitely Increasing
DM28	Suppose you hold [Company] stock. How would you change your holdings of [Company] stock?	Significantly Decrease	Neither Decrease Nor Increase	Significantly Increase
DM29	Indicate how attractive [Company]'s stock is as an investment.	Very Unattractive	Neither Unattractive Nor Attractive	Very Attractive
DM30	What is the likelihood you would consider [Company]'s stock as a potential investment?	Very Low	About Average	Very High
DM31	How do you think that potential investors would evaluate [Company]'s stock in terms of its desirability as a potential investment?	Very Undesirable	About Average	Very Desirable
DM32	Rate the potential for [Company]'s stock price to appreciate over the next 12 months.	Very Low Potential	About Average Potential	Very High Potential
DM33	How willing are you to invest in [Company] stock?	Very Unwilling	Neither Unwilling Nor Willing	Very Willing

TABLE 1 (CONTINUED)

Ref.	**Measure**	**1**	**Mid**	**7**
DM34	Do you view [Company]'s stock as a more favorable or less favorable investment?	Significantly Less Favorable	Neither Less Nor More Favorable	Significantly More Favorable
DM35	Are your feelings towards [Company]'s stock as a potential investment generally more positive or more negative?	Significantly Negative	Neither Negative Nor Positive	Significantly Positive
DM36	Rate the extent to which you agree with the following statement: "I believe buying [Company]'s stock would be a good investment choice."	Strongly Disagree	Neither Disagree Nor Agree	Strongly Agree
DM37	My general perceptions of [Company]'s stock as a potential investment are _______.	Significantly Negative	Neutral	Significantly Positive
DM38	Do you view [Company]'s stock more favorable or less favorable in general?	Significantly Less Favorable	Neither Less Nor More Favorable	Significantly More Favorable

Table 1 lists the 38 dependent measures and associated response scales we extracted from our survey of the investor judgment literature. We first identified all dependent measures used in prior literature, and then combined measures that were nearly identical or redundant. In addition, we added several measures that elicit more general perceptions about the firm or future firm prospects. Further, we adapt all scales to be 7-point Likert-style response scales and use qualitative labels for all measures (e.g., Very low to Very high; Very weak to Very strong; etc.).

TABLE 2. EXPLORATORY FACTOR ANALYSIS PARTICIPANT DEMOGRAPHICS

Gender:	
Male	52%
Female	47%
Other/Prefer not to say	1%
Age Percentiles:	
1%:	20 years
25%:	31
Median:	38
Mean:	41
75%	49
99%	74
Education:	
Did not graduate high school	1%
High school or equivalent	8%
Vocational/technical school	5%
Some college	20%
Bachelor's degree	49%
Master's degree	14%
Doctoral degree	1%
Professional degree (MD, JD, etc.)	1%
Other (e.g., associate degree)	1%
Number of participants investing or planning to invest in various asset types:	
None	9%
Individual company stocks	57%
Mutual funds	48%
Index funds	25%
401(k) plans	65%
Government bonds	21%
Corporate bonds	10%
Total value of investment in stocks, bonds, mutual funds, IRA's, 401(k) plans, and the like:	
$0 - $4,999	32%
$5,000 - $9,999	10%
$10,000 - $19,999	9%
$20,000 - $29,999	8%
$30,000 - $49,999	9%
$50,000 or more	32%

Table 2 provides demographic information for the Amazon Mechanical Turk participants recruited for the exploratory factor analysis. All percentages are based on the total number of completed responses, which is 999.

TABLE 3. CORRELATION MATRIX FROM EXPLORATORY FACTOR ANALYIS

	DM1	**DM2**	**DM3**	**DM4**	**DM5**	**DM6**	**DM7**	**DM8**	**DM9**	**DM10**	**DM11**	**DM12**	**DM13**	**DM14**	**DM15**	**DM16**	**DM17**	**DM18**	**DM19**
DM1	1.00																		
DM2	0.85	1.00																	
DM3	0.86	0.85	1.00																
DM4	0.84	0.82	0.83	1.00															
DM5	0.88	0.85	0.88	0.83	1.00														
DM6	0.87	0.85	0.88	0.84	0.88	1.00													
DM7	0.85	0.82	0.86	0.81	0.87	0.86	1.00												
DM8	-0.40	-0.38	-0.35	-0.42	-0.37	-0.35	-0.41	1.00											
DM9	-0.44	-0.45	-0.41	-0.46	-0.43	-0.42	-0.43	0.74	1.00										
DM10	-0.38	-0.37	-0.36	-0.40	-0.36	-0.36	-0.40	0.83	0.72	1.00									
DM11	-0.44	-0.44	-0.40	-0.44	-0.42	-0.41	-0.43	0.63	0.55	0.61	1.00								
DM12	-0.29	-0.29	-0.27	-0.33	-0.28	-0.29	-0.31	0.71	0.62	0.72	0.52	1.00							
DM13	0.71	0.75	0.75	0.70	0.72	0.72	0.69	-0.19	-0.27	-0.17	-0.28	-0.14	1.00						
DM14	-0.46	-0.46	-0.42	-0.46	-0.46	-0.45	-0.46	0.61	0.67	0.61	0.50	0.50	-0.23	1.00					
DM15	0.79	0.77	0.77	0.76	0.80	0.77	0.77	-0.40	-0.42	-0.39	-0.43	-0.33	0.64	-0.44	1.00				
DM16	0.83	0.80	0.81	0.81	0.82	0.82	0.83	-0.40	-0.45	-0.40	-0.42	-0.32	0.71	-0.48	0.77	1.00			
DM17	0.77	0.74	0.72	0.76	0.75	0.73	0.73	-0.42	-0.42	-0.41	-0.49	-0.34	0.61	-0.43	0.78	0.73	1.00		
DM18	0.73	0.71	0.69	0.72	0.73	0.69	0.72	-0.44	-0.41	-0.43	-0.50	-0.34	0.59	-0.43	0.76	0.71	0.78	1.00	
DM19	0.77	0.75	0.73	0.74	0.76	0.72	0.73	-0.40	-0.40	-0.39	-0.48	-0.31	0.61	-0.42	0.79	0.73	0.87	0.78	1.00
DM20	0.79	0.79	0.81	0.77	0.81	0.79	0.79	-0.38	-0.41	-0.37	-0.40	-0.29	0.68	-0.40	0.79	0.78	0.72	0.72	0.74
DM21	0.81	0.80	0.82	0.78	0.82	0.81	0.80	-0.39	-0.42	-0.37	-0.42	-0.30	0.68	-0.43	0.83	0.80	0.76	0.73	0.77
DM22	0.82	0.81	0.83	0.80	0.83	0.81	0.81	-0.40	-0.43	-0.40	-0.43	-0.31	0.69	-0.43	0.80	0.80	0.76	0.74	0.76
DM23	0.80	0.79	0.80	0.78	0.80	0.78	0.78	-0.40	-0.43	-0.38	-0.45	-0.31	0.67	-0.44	0.83	0.78	0.79	0.77	0.79
DM24	0.61	0.62	0.58	0.62	0.60	0.57	0.57	-0.31	-0.31	-0.31	-0.44	-0.20	0.51	-0.35	0.61	0.56	0.69	0.64	0.69
DM25	0.69	0.70	0.66	0.70	0.68	0.66	0.68	-0.44	-0.44	-0.44	-0.48	-0.36	0.58	-0.44	0.71	0.67	0.74	0.71	0.75
DM26	0.48	0.47	0.44	0.46	0.45	0.42	0.46	-0.38	-0.31	-0.37	-0.43	-0.30	0.38	-0.34	0.51	0.46	0.54	0.54	0.54
DM27	0.74	0.71	0.71	0.73	0.72	0.71	0.74	-0.48	-0.55	-0.48	-0.47	-0.39	0.61	-0.57	0.66	0.75	0.66	0.66	0.65
DM28	0.63	0.64	0.60	0.63	0.61	0.60	0.60	-0.36	-0.36	-0.37	-0.43	-0.27	0.54	-0.39	0.65	0.60	0.73	0.68	0.73
DM29	0.75	0.76	0.72	0.73	0.73	0.72	0.73	-0.46	-0.45	-0.45	-0.53	-0.34	0.61	-0.46	0.74	0.70	0.82	0.78	0.83
DM30	0.68	0.69	0.66	0.68	0.68	0.65	0.64	-0.40	-0.40	-0.41	-0.54	-0.31	0.57	-0.42	0.69	0.64	0.80	0.73	0.79
DM31	0.78	0.80	0.79	0.77	0.80	0.78	0.78	-0.43	-0.45	-0.43	-0.46	-0.37	0.69	-0.44	0.78	0.77	0.74	0.74	0.76
DM32	0.81	0.80	0.83	0.79	0.80	0.81	0.79	-0.35	-0.42	-0.34	-0.40	-0.28	0.73	-0.44	0.75	0.80	0.71	0.68	0.73
DM33	0.66	0.68	0.62	0.68	0.64	0.62	0.63	-0.43	-0.42	-0.42	-0.55	-0.34	0.56	-0.44	0.67	0.63	0.78	0.72	0.76
DM34	0.71	0.70	0.67	0.71	0.69	0.66	0.68	-0.45	-0.44	-0.46	-0.55	-0.36	0.58	-0.47	0.71	0.66	0.81	0.75	0.79
DM35	0.67	0.68	0.63	0.68	0.65	0.62	0.64	-0.46	-0.44	-0.45	-0.55	-0.38	0.55	-0.45	0.67	0.62	0.79	0.75	0.77
DM36	0.74	0.75	0.70	0.73	0.71	0.69	0.71	-0.48	-0.49	-0.46	-0.55	-0.37	0.61	-0.49	0.73	0.69	0.81	0.77	0.80
DM37	0.71	0.72	0.67	0.70	0.68	0.66	0.68	-0.46	-0.46	-0.45	-0.56	-0.35	0.57	-0.47	0.71	0.67	0.81	0.78	0.80
DM38	0.67	0.68	0.64	0.68	0.66	0.63	0.65	-0.45	-0.42	-0.45	-0.55	-0.35	0.55	-0.45	0.69	0.63	0.80	0.76	0.78

TABLE 3 (CONTINUED)

	DM20	**DM21**	**DM22**	**DM23**	**DM24**	**DM25**	**DM26**	**DM27**	**DM28**	**DM29**	**DM30**	**DM31**	**DM32**	**DM33**	**DM34**	**DM35**	**DM36**	**DM37**	**DM38**
DM20	1.00																		
DM21	0.86	1.00																	
DM22	0.86	0.88	1.00																
DM23	0.81	0.84	0.84	1.00															
DM24	0.56	0.60	0.60	0.62	1.00														
DM25	0.65	0.67	0.67	0.70	0.67	1.00													
DM26	0.44	0.48	0.48	0.49	0.51	0.57	1.00												
DM27	0.68	0.68	0.69	0.68	0.54	0.66	0.49	1.00											
DM28	0.58	0.62	0.62	0.64	0.72	0.76	0.53	0.61	1.00										
DM29	0.70	0.73	0.73	0.77	0.73	0.79	0.58	0.67	0.77	1.00									
DM30	0.63	0.67	0.67	0.70	0.78	0.77	0.59	0.61	0.80	0.86	1.00								
DM31	0.82	0.82	0.82	0.79	0.61	0.70	0.49	0.70	0.64	0.77	0.72	1.00							
DM32	0.77	0.79	0.79	0.79	0.57	0.66	0.46	0.72	0.58	0.70	0.65	0.77	1.00						
DM33	0.61	0.64	0.64	0.68	0.79	0.78	0.60	0.62	0.80	0.84	0.90	0.69	0.64	1.00					
DM34	0.65	0.68	0.68	0.73	0.74	0.77	0.58	0.65	0.78	0.88	0.87	0.73	0.66	0.88	1.00				
DM35	0.61	0.64	0.64	0.68	0.75	0.78	0.59	0.63	0.78	0.86	0.87	0.70	0.62	0.88	0.89	1.00			
DM36	0.70	0.71	0.72	0.75	0.73	0.80	0.59	0.69	0.77	0.87	0.86	0.76	0.70	0.86	0.87	0.85	1.00		
DM37	0.66	0.69	0.68	0.73	0.75	0.78	0.59	0.65	0.77	0.88	0.88	0.74	0.66	0.88	0.89	0.91	0.87	1.00	
DM38	0.63	0.66	0.65	0.70	0.77	0.77	0.61	0.63	0.79	0.85	0.88	0.71	0.62	0.89	0.90	0.90	0.85	0.90	1.00

Table 3 presents the correlations among the dependent measures based on responses received from 999 Amazon Mechanical Turk workers who participated in the exploratory factor analysis.

TABLE 4. EXPLORATORY FACTOR ANALYIS THREE-FACTOR MODEL

PANEL A. Factor loadings, variance explained, and factor correlations

Measure	Factor1	Factor2	Factor3	Communality	Uniqueness	Item Complexity
DM6	0.99			0.85	0.15	1.0
DM3	0.98			0.86	0.14	1.0
DM5	0.95			0.87	0.13	1.0
DM7	0.91			0.83	0.17	1.0
DM20	0.89			0.78	0.22	1.0
DM1	0.89			0.85	0.15	1.0
DM16	0.89			0.80	0.20	1.0
DM22	0.88			0.82	0.18	1.0
DM32	0.87			0.78	0.22	1.0
DM21	0.87			0.82	0.18	1.0
DM2	0.83			0.83	0.17	1.0
DM13	0.82			0.64	0.36	1.1
DM4	0.80			0.80	0.20	1.0
DM23	0.74			0.79	0.21	1.1
DM15	0.71			0.76	0.24	1.1
DM31	0.68			0.78	0.22	1.2
DM27	0.61			0.66	0.34	1.3
DM33		0.99		0.88	0.12	1.0
DM38		0.98		0.90	0.10	1.0
DM35		0.97		0.89	0.11	1.0
DM30		0.95		0.87	0.13	1.0
DM37		0.91		0.90	0.10	1.0
DM34		0.90		0.88	0.12	1.0
DM24		0.82		0.67	0.33	1.0
DM28		0.82		0.71	0.29	1.0
DM36		0.75		0.86	0.14	1.1
DM29		0.74		0.86	0.14	1.2
DM25		0.61		0.72	0.28	1.3
DM26		0.57		0.41	0.59	1.1
DM17	0.38	0.56		0.79	0.21	1.8
DM19	0.42	0.53		0.79	0.21	1.9
DM18	0.40	0.46		0.72	0.28	2.0
DM8			0.91	0.83	0.17	1.0
DM10			0.90	0.81	0.19	1.0
DM12			0.80	0.61	0.39	1.0
DM9			0.77	0.68	0.32	1.1
DM14			0.59	0.52	0.48	1.2
DM11		-0.36	0.52	0.53	0.47	1.8

	Factor1	**Factor2**	**Factor3**
SS loadings	14.24	10.99	4.11
Proportion Var	37%	29%	11%
Cumulative Var	37%	66%	77%

Factor correlations

	Factor1	Factor2	Factor3
Factor1	1.00		
Factor2	0.78	1.00	
Factor3	-0.45	-0.49	1.00

Mean item complexity = 1.1

TABLE 4 (CONTINUED)

PANEL B. Measurement Model	
Factor 1 =~	DM6 + DM3 + DM5 + DM7 + DM20 + DM1 + DM16 + DM22 + DM32 + DM21 + DM2 + DM13 + DM4 + DM23 + DM15 + DM31 + DM27
Factor 2 =~	DM33 + DM38 + DM35 + DM30 + DM37 + DM34 + DM24 + DM28 + DM36 + DM29 + DM25 + DM26
Factor 3 =~	DM8 + DM10 + DM12 + DM9 + DM14
Factor 1 ~	Future Earnings Manipulation + ESG Manipulation + Risk Manipulation
Factor 2 ~	Future Earnings Manipulation + ESG Manipulation + Risk Manipulation
Factor 3 ~	Future Earnings Manipulation + ESG Manipulation + Risk Manipulation
Factor 1 ~~	Factor 2
Factor 1 ~~	Factor 3
Factor 2 ~~	Factor 3

PANEL C. SEM Coefficient Estimates

	Factor 1	Factor 2	Factor 3
Future Earnings Manipulation	0.62***	0.38***	-0.18***
ESG Manipulation	0.26***	0.57***	-0.32***
Risk Manipulation	-0.24***	-0.22***	0.53***

Table 4, Panel A presents the exploratory factor analysis output obtained by using the ‘fa’ function from the ‘lavaan’ package in the R statistical software. We use maximum likelihood estimation to estimate the factor loadings and use a direct oblimin oblique rotation to increase the interpretability of the factor solution. Only factor loadings above 0.30 are displayed. DMs are listed within each factor by their factor loading. Sum of square loadings are presented for each factor, as well as the total variance among the DMs that is explained by each factor. Finally, the correlations between each factor are presented, as well as the mean item complexity score for the entire dataset. See Appendix A for definitions of item complexity, uniqueness, and communality.
Table 4, Panel B presents the measurement model we specify in the R statistical software to assess the influence of each experimental manipulation on each factor. First, each DM is assigned to each factor based on the factor loadings from Panel A. DM11, DM17, DM18, and DM19 are excluded because they cross-load onto multiple factors. Next, Factors 1, 2, and 3 are regressed on an indicator variable for each experimental manipulation. Finally, Factors, 1, 2, and 3 are allowed to correlate with one another.
Table 4, Panel C presents the SEM coefficient estimates from the regressions of Factors 1, 2, and 3 on each experimental manipulation.
*** indicates that the coefficient is statistically significant at the 1% level.

TABLE 5. INTERPRETATION OF EXPLORATORY FACTOR ANALYSIS SOLUTION

Panel A - Factor 1: Expectations regarding future firm performance and value	
Ref.	**Measure**
DM6	You **forecast [Company]'s earnings growth** rate for the next year to be _______.
DM3	[Company]**'s earnings potential** over the next year is _______.
DM5	You **forecast [Company]'s earnings** for the next year to be _______.
DM7	You **forecast [Company]'s cash flows** for the next year to be _______.
DM20	I believe that other stock market participants would **value [Company]'s stock** at a _______.
DM1	[Company]'s **future earnings performance** will be _______.
DM16	What do you **expect [Company]'s stock price** per share to be next year?
DM22	How do you believe investors would **perceive the value of [Company]** stock?
DM32	Rate the **potential for [Company]'s stock price to appreciate** over the next 12 months.
DM21	What **common stock valuation** do you think potential investors would place on [Company] stock?
DM2	Rate the extent to which you agree with the following statement: "[Company]'s **earnings performance will be** strong in the near future."
DM13	Rate the extent to which you agree with the following statement: "Over the next 12 months, the **stock price could increase significantly**."
DM4	How much do you think [Company]'s **earnings will grow** in the future?
DM23	Provide a judgment about the **appropriate valuation of [Company]** stock.
DM15	What do you believe is an **appropriate common stock valuation for [Company]**?
DM31	How do you think that potential investors would evaluate **[Company]'s stock in terms of its desirability** as a potential investment?
DM27	Assume you were placing a $10,000 bet on **[Company]'s future stock price**. Would you bet on [Company]'s stock price increasing or decreasing over the next year?

TABLE 5 (CONTINUED)

Panel B - Factor 2: Holistic perceptions of the firm	
Ref.	**Measure**
DM33	How **willing are you to invest** in [Company] stock?
DM38	Do you view [Company]'s stock more favorable or less **favorable in general**?
DM35	Are **your feelings towards [Company]'s stock** as a potential investment generally more positive or more negative?
DM30	What is the likelihood you would **consider [Company]'s stock as a potential investment**?
DM37	My **general perceptions of [Company]'s stock** as a potential investment are _______.
DM34	Do you view [Company]'s stock as a **more favorable or less favorable investment**?
DM24	How much of a $10,000 bonus **would you invest in [Company]'s stock**?
DM28	Suppose you hold [Company] stock. How **would you change your holdings** of [Company] stock?
DM36	Rate the extent to which you agree with the following statement: "**I believe buying [Company]'s stock would be a good investment choice**."
DM29	Indicate **how attractive [Company]'s stock is** as an investment.
DM25	**Provide a buy/sell/hold recommendation** for [Company]'s stock.
DM26	What would be **your preferred investment position** in [Company]'s stock?

TABLE 5 (CONTINUED)

Panel C - Factor 3: Evaluations of the risk associated with investing in the firm	
Ref.	**Measure**
DM8	**How risky do you consider an investment** in [Company] stock to be?
DM10	**Rate the risk associated with investing** in [Company] stock.
DM12	**The investment risk of [Company]** in the context of a diversified portfolio is **_______**.
DM9	What is **the risk of [Company]** experiencing a moderate stock price decline within the next year?
DM14	Rate the extent to which you agree with the following statement: "Over the next 12 months, **the stock price could decrease significantly**."

Table 5, Panel A presents the dependent measures that load onto Factor 1: Expectations regarding future firm performance and value.
Table 5, Panel B presents the dependent measures that load onto Factor 2: Holistic perceptions of the firm.
Table 5, Panel C presents the dependent measures that load onto Factor 3: Evaluations of the risk associated with investing in the firm.
Note: In each panel, dependent measures are listed in descending order of factor loadings obtained from the exploratory factor analysis. DM11, DM17, DM18, and DM19 all cross-load onto multiple factors, and are not listed. *Please note that we caution against simply selecting dependent measures from the table based on their factor loadings alone.*

TABLE 6. CONFIRMATORY FACTOR ANALYSIS PARTICIPANT DEMOGRAPHICS

Gender:	
Male	52%
Female	47%
Other/Prefer not to say	1%
Age Percentiles:	
1%:	20 years
25%:	30
50%:	37
75%	48
99%	73
Education:	
Did not graduate high school	< 1%
High school or equivalent	10%
Vocational/technical school	4%
Some college	21%
Bachelor’s degree	45%
Master’s degree	14%
Doctoral degree	2%
Professional degree (MD, JD, etc.)	3%
Other (e.g., associate degree)	< 1%
Number of participants investing or planning to invest in various asset types:	
None	9%
Individual company stocks	61%
Mutual funds	49%
Index funds	29%
401(k) plans	64%
Government bonds	17%
Corporate bonds	8%
Total value of investment in stocks, bonds, mutual funds, IRA’s, 401(k) plans, and the like:	
$0 - $4,999	29%
$5,000 - $9,999	9%
$10,000 - $19,999	13%
$20,000 - $29,999	8%
$30,000 - $49,999	9%
$50,000 or more	31%

Table 6 provides demographic information for the Amazon Mechanical Turk participants recruited for the confirmatory factor analysis. All percentages are based on the total number of completed responses, which is 998.

TABLE 7. CONFIRMATORY FACTOR ANALYSIS OUTPUT AND MODEL COMPARISON

PANEL A. Three-factor measurement model and output

Factor 1 =~	DM6 + DM3 + DM5 + DM7 + DM20 + DM1 + DM16 + DM22 + DM32 + DM21 + DM2 + DM13 + DM4 + DM23 + DM15 + DM31 + DM27
Factor 2 =~	DM33 + DM38 + DM35 + DM30 + DM37 + DM34 + DM24 + DM28 + DM36 + DM29 + DM25 + DM26
Factor 3 =~	DM8 + DM10 + DM12 + DM9 + DM14
Factor 1 ~~	Factor 2
Factor 1 ~~	Factor 3
Factor 2 ~~	Factor 3

Measure	Factor1	Factor2	Factor3
DM6	0.92		
DM3	0.92		
DM5	0.92		
DM7	0.90		
DM20	0.86		
DM1	0.91		
DM16	0.89		
DM22	0.87		
DM32	0.88		
DM21	0.87		
DM2	0.90		
DM13	0.78		
DM4	0.87		
DM23	0.86		
DM15	0.85		
DM31	0.89		
DM27	0.84		
DM33		0.91	
DM38		0.93	
DM35		0.94	
DM30		0.90	
DM37		0.93	
DM34		0.93	
DM24		0.74	
DM28		0.80	
DM36		0.92	
DM29		0.93	
DM25		0.83	
DM26		0.62	
DM8			0.88
DM10			0.87
DM12			0.73
DM9			0.78
DM14			0.69

Factor Covariance

Factor 1 ~~ Factor 2	0.96
Factor 1 ~~ Factor 3	-0.66
Factor 1 ~~ Factor 3	-0.71

TABLE 7 (CONTINUED)

PANEL B. Fit statistics
χ^2 = 2,359.67 (*df* = 524; *p*-value = < 0.001)
CFI = 0.958
TLI = 0.955
RMSEA = 0.059, RMSEA lower = 0.057, RMSEA upper = 0.062 (H_0: RMSEA $\leq$ 0.05, p < 0.001)
SRMR = 0.033

Table 7, Panel A presents the three-factor measurement model we specify in the R statistical software to conduct the confirmatory factor analysis to assess model fit. First, each DM is assigned to each factor based on the factor loadings from the exploratory factor analysis. DM11, DM17, DM18, and DM19 are excluded because they cross-load onto multiple factors. Next, Factors 1, 2, and 3 are allowed to correlate with one another. The next section shows the standardized coefficient estimates for each DM. Finally, the last section shows the factor covariances.

Table 7, Panel B presents the model fit statistics for the three-factor measurement model specified in Panel A. See Appendix A for definitions of model chi-square, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR).