Reducing Underreporting by Aggregating Budgeted Time

Kimberly M. Ikuta

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Frank Hodge, Chair

Dawn Matsumoto

John Miyamoto

Program Authorized to Offer Degree:

Business Administration

University of Washington

Abstract

Reducing Underreporting by Aggregating Budgeted Time

Kimberly Midori Ikuta

Chair of the Supervisory Committee:
Dr. Frank Hodge
Business Administration

Underreporting, or reporting fewer hours than actually worked, is a prevalent behavior among auditors at all levels. Underreporting can result in negative consequences such as tight budgets and reductions in future audit quality. In this paper, I propose a low cost reporting procedure that reduces underreporting. Using an experiment, I document that individuals with incentives to underreport report more accurately when reporting aggregated—relative to disaggregated—time. In contrast, when individuals do not face underreporting incentives aggregation does not influence reporting. Building on mental accounting theory, I also provide evidence suggesting that the utility loss individuals' mitigate by underreporting mediates the relation between the level of aggregation, underreporting incentives, and the degree of underreporting. In a second experiment, I document another reporting procedure—percentage reporting—that firms can use in concert with aggregation to mitigate the loss of data richness that results from aggregation. This study provides important insights to audit firms, partners, managers, and regulators who rely on audit hours for budgets, measures of staff efficiency, and measures of audit quality.

I. INTRODUCTION

Underreporting, or reporting fewer hours than actually worked, is a prevalent behavior among auditors at all levels (Taylor, Curtis, and Chui 2012). Auditors often underreport their hours in response to both manager pressures and the desire to appear more efficient (Agoglia, Hatfield, and Lambert 2015). Underreporting threatens audit quality by perpetuating tight budgets, which may cause auditors to perform inferior procedures (Kelley and Margheim 1990). Recently, some audit firms have begun publicly disclosing measures of audit quality that use audit hours as an input, and the Public Company Accounting Oversight Board (PCAOB) is considering using similar audit hour metrics to evaluate audit firms' audit quality (PwC 2015; PCAOB 2015). Faced with these matters, both regulators and audit firms have expressed concerns about auditor underreporting (HOL 2010; POB 2000; McNair 1991). Despite firms' attempts to curtail underreporting—e.g., by explicitly prohibiting the practice or reducing their emphasis on budgets (Buchheit, Pasewark and Strawser 2003)—underreporting continues to exist.

In this paper, I propose and experimentally test a reporting procedure that causes auditors to report more accurate audit hours—despite incentives to underreport. Specifically, I investigate whether firms can reduce underreporting by asking auditors to report their time aggregated as one number (e.g., reporting the total number of hours worked) as opposed to disaggregated as separate numbers (e.g., reporting the number of hours spent working on each individual audit task). I also investigate the impact of aggregation on auditor underreporting when auditors do or do not have incentives to underreport. In doing so, I build on mental accounting theory and generate new predictions about when misreporting, in general, is more or less prevalent.

Examining a reporting procedure that reduces auditor underreporting is important for several reasons. First, reducing underreporting could improve audit firms' internal decision-making because many important decisions are based on metrics that use reported audit hours. For example, audit hours are often used to create budgets that determine the audit fees a firm will charge. Additionally, audit hours often serve as a benchmark for partners and managers when evaluating their audit team's performance (McNair 1991).

Second, prior literature provides evidence that tight audit budgets result in several negative consequences for audit quality, such as superficial reviews, premature sign-off of workpapers, accepting weak client explanations to auditor inquiries, and insufficient research of technical issues (e.g., Kelley and Margheim 1990). Because underreporting leads to tighter future budgets, reducing underreporting could help audit firms begin the process of unwinding the tight budgets that have accumulated over time, thereby improving future audit quality.

Third, presumably because audit hours serve as a rough proxy for both the effort spent on an area and the quality of the audit performed over that area, audit firms have recently started using audit hours in their public disclosures of audit quality (PwC 2015). Similarly, the PCAOB has proposed using audit hours to measure and evaluate firms' audit quality (PCAOB 2015). Thus, audit hour accuracy has become even more important in the current audit environment than in the past. Given the current focus on measuring audit quality, my results should be of interest to firms and regulators as I propose a technique to increase audit hour accuracy.

I develop my predictions using the theory of mental accounting. Mental accounting predicts individuals' preferences for aggregated or disaggregated outcomes by demonstrating how aggregation and disaggregation result in different utility gains or losses (Thaler 1985). Mental accounting takes outcomes as given and does not make predictions about how individuals

might manipulate their outcomes. In a reporting setting, preparers can manipulate reported information, which results in differences between actual and reported outcomes. By combining the ability to manipulate reported outcomes with the theory of mental accounting, I generate predictions about when there will be more or less misreporting. Specifically, I predict that when there are incentives to underreport, individuals will underreport more when reporting disaggregated (relative to aggregated) hours because the disaggregated format allows the to mitigate a greater utility loss. In contrast, without incentives to underreport, the resulting utility loss mitigated by underreporting is held constant across aggregation formats, and disaggregation does not impact the degree of misreporting.

In this study, I conduct two experiments. My first experiment examines whether aggregating benchmarks reduces underreporting when incentives to underreport are present. I use a 2 × 2 (aggregated vs. disaggregated benchmarks × incentives present vs. absent) between-participants design. Participants begin by learning that they worked on eight tasks, and that they must self-report their hours worked to their manager. For my incentive manipulation, I tell participants that their time either is or is not important for performance reviews and obtaining promotion opportunities. Next, all participants learn the "true" number of hours worked, which is over the budgeted number of hours. I then manipulate whether participants report their time in an aggregated format (i.e., reporting total hours worked compared to a "total hours" benchmark) or a disaggregated format (i.e., reporting hours worked in eight separate tasks compared to a benchmark for each task). I measure the degree of underreporting as the difference between the true total number of hours and the reported total number of hours.

Although aggregation may increase accuracy, it is not costless. When the time spent on various audit tasks is aggregated into a single number, audit firms lose the rich data that helps

create detailed budgets. To address this trade-off, my second experiment investigates whether audit firms can achieve both data richness and accuracy by using percentage reporting with aggregation. Specifically, I test whether asking participants to provide a percentage of time spent on each task, after reporting aggregated hours, impacts the level of underreporting compared to participants who only report aggregated time. I expect that percentages may provide some data richness while still maintaining the benefits of aggregation because individuals respond to percentages differently than they do to raw numbers (e.g., Nelson and Rupar 2015).

Experiments are the ideal methodology for examining my research questions for a number of reasons. In the natural setting, the true number of hours worked is observable only by the auditor reporting the hours, therefore there is no precise way to measure underreporting using other methods (Poneman 1992; McNair 1991). In my experiment, I can observe the true number of hours because I provide participants with a "true" number, which I use to measure the degree of underreporting. Additionally, in an experiment I can hold constant external factors that impact the number of hours reported (e.g., task complexity and manager pressures). These factors are difficult to measure and control for using other methods. Finally, through an experiment I can investigate the process or *why* auditor underreporting varies with aggregation.

My results provide a number of important insights. I document that participants underreport more when they have underreporting incentives and report disaggregated time compared to when participants have underreporting incentives and report aggregated time or have no underreporting incentives. I also find that aggregation does not significantly impact reporting when auditors do not face underreporting incentives. In a mediation analysis, I observe that the loss of utility participants mitigate by underreporting mediates the relationship between

aggregation, incentives, and the degree of underreporting. Lastly, I demonstrate that audit firms can obtain accuracy and data richness by using aggregation along with percentage reporting.

This study has both practical and theoretical contributions. First, my study contributes to an extensive literature examining auditor underreporting. The prior literature on underreporting focuses on its determinants (e.g., Agoglia et al. 2015) and its consequences (e.g., Donnelly, Quirin, and O'Bryan 2003). These prior studies establish whether and why auditor underreporting occurs; however, they do not document a clear solution. Reducing underreporting is challenging as many fixes are costly (e.g., increased monitoring of audit staff time). Therefore, I extend this literature by testing a low-cost, implementable reporting procedure that reduces underreporting.

This study also contributes to the psychology literature. Specifically, it builds on the theory of mental accounting by incorporating an individual's ability to manipulate outcomes. Currently, mental accounting predicts *preferences* for aggregation or disaggregation of multiple outcomes. In contrast, this paper hypothesizes *actions* individuals take to arrive at a particular outcome when these outcomes are aggregated versus disaggregated.

Finally, this study also contributes to the accounting literature on disaggregation. The theory in this study generalizes to settings where individuals report financial information in varying levels of disaggregation (e.g., financial statement line item reporting, or segment reporting), adding to papers that examine manager and analyst preferences for aggregated or disaggregated reporting (Bonner, Clor-Proell, and Koonce 2014; Fennema and Koonce 2010). The most closely related study to mine is Chen Rennekamp, and Zhou (2015), which investigates disaggregation and reporting incentives in the management forecast setting. Specifically, Chen et al. (2015) document that because ambiguity exists in the management forecast setting, when

managers face incentives they unintentionally bias their forecasts more when they report in a disaggregated format versus an aggregated format. My study is distinct from Chen et al. (2015) as I examine a setting that lacks ambiguity (i.e., there is a true number to report). As such any reporting bias I document is *intentional* bias, and therefore my theory is distinct for why reporting bias varies with disaggregation.

In the next section, I discuss background and develop my predictions. In Sections III and IV, I describe my research design and results for Experiments 1 and 2, respectively. I conclude the paper in section V, where I discuss the study's implications and limitations.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

Reporting Audit Hours

Reporting the amount of time one spends auditing is vital to the audit process. An auditor can report his or her number of hours worked in varying degrees of detail depending on the firm's, manager's, or partner's requirements. For example, auditors might report the amount of time spent working on a client in aggregate, by audit task, by week, by financial statement area audited, by day, or any combination of these. The exact number of hours an auditor spends working on each task is observable only by the auditor performing the work. Consequently, the only practical way for audit firms, partners, and managers to obtain an estimate of the time spent working on a particular client or audit area is to require staff to self-report their hours (McNair 1991). However, given that audit firms, partners, and managers cannot determine whether an auditor is misreporting time, if auditors face incentives to underreport, then audit hour accuracy may suffer.

Audit hour accuracy is important because auditors and audit firms use reported hours for a number of important metrics (for an overview see McNair 1991). First, partners often use the

reported hours for a client in the current year to determine the fee to charge and the employee resources to use in the following year. Second, firms compare an audit's reported hours to the audit's fee to determine realization rates, a statistic used to evaluate managers' and partners' revenue generation. Finally, firms, partners, and managers use audit hours to evaluate individual auditors to determine if individual auditors are working efficiently on their assigned jobs.

Audit Firms' and the PCAOB's use of Audit Hours in Measures of Audit Quality

In recent years, the Big 4 accounting firms have released annual audit quality reports and transparency reports, which are outlets the firms use to discuss firm performance metrics and goals. The firms spend a significant amount of time detailing their commitment to audit quality, and some firms report metrics as evidence of their commitment to audit quality (Deloitte 2016; PwC 2015; KPMG 2015; EY 2015). Several of these audit quality metrics use audit hours as an input (e.g., hours worked in excess of 40 hours per week, and percentage of specialist hours) (PwC 2015).

Audit hours are not only important for audit firms, audit hours have also recently entered the PCAOB's discussion around audit quality. On July 1, 2015, the PCAOB issued a Concept Release proposing 28 audit quality indicators that the PCAOB could use to assess the quality of audit engagements and audit firms. Eight of the 28 audit quality indicators involve some reliance on actual audit hours. Examples include staffing leverage (e.g., ratio of experienced personnel hours to the hours of the auditors they oversee), manager and staff workload (e.g., actual hours compared to the budget), and allocation of audit hours to phases of the audit (e.g., audit hours charged to each part of the audit) (PCAOB 2015). Given the role that audit hours play in assessing audit quality, both in the current and future audit environments, audit firms and standard setters may find it helpful to learn that requiring audit hour data at greater levels of

disaggregation leads to greater degrees of underreporting, making their audit quality indicators less accurate.

Underreporting

Underreporting, or reporting fewer hours than one actually worked, is a pervasive phenomenon within the audit profession. Indeed, prior research by Lightner (1982) and Lightner, Adams, and Lightner (1982) documents that approximately 65 percent of auditors admit to underreporting, and that underreporting occurs at all levels within an audit firm. Lightner, Leisering, and Winters (1983) conduct a survey and find that 40 percent of respondents disclose that they underreport at least 5 percent of the hours they work. Recent studies observe that even with firms' efforts to reduce underreporting the phenomenon still exists (Agoglia et al. 2015; Taylor et al. 2012). The prevalence and degree of underreporting among auditors at all levels demonstrates the significance of this phenomenon.

Underreporting leads to several negative consequences that impact audit firms, auditors, audit quality, and measures of audit quality. First, underreporting perpetuates artificially tight audit budgets. Auditors use underreported audit hours to create future budgets that are then tighter than they would be with actual hours as an input. To make matters worse, these tight budgets then impact next year's reported audit hours, which perpetuates the cycle. Prior studies document that tight budgets reduce audit quality because auditors lack time to gather sufficient evidence, leading to the omission of significant audit findings and documentation of procedures that were never performed (Coram, Ng, and Woodliff 2004; Donnelly et al. 2003; Otley and Pierce 1996). Thus, by perpetuating tight budgets, underreporting leads to reductions in future audit quality.[1] Second, underreporting results in an unequal or unfair division of benefits, as

[1] Importantly, not all individuals underreport and not all audit engagements result in underreporting. For example, if a budget is ample and auditors are able to complete their tasks using only the budgeted number of hours, then

some people are willing to underreport and others are not. For example, a study by Agoglia et al. (2015) documents that managers tend to select underreporters more often to work on future audits, and sometimes even give underreporters higher performance evaluations. Third, if audit firms and the PCAOB use audit hours in their measures of audit quality, these measures may be inaccurate as a result of misreported hours.

Underreporting audit hours occurs because auditors have the opportunity to underreport and because there are incentives to underreport.[2] Self-reporting time results in an opportunity for underreporting because there is no practical way for one's superiors to check the exact accuracy of the reported information. Auditors face many pressures or incentives to underreport. For example, auditors frequently encounter tight budgets, which exert pressure to be unrealistically efficient (Kelley and Margheim 1990; Margheim and Pany 1986). Also, auditors can face explicit or implicit pressure from their superiors to underreport when they are not meeting the budget (Agoglia et al. 2015; Taylor et al. 2012; Lightner et al. 1982). Additionally, auditors can reap benefits from underreporting such as the ability to appear more valuable or efficient than their peers, which could result in better performance evaluations and promotion opportunities (Sweeney and Pierce 2006). All of these incentives lead to the pervasive underreporting that exists in the audit profession.

It is important to note that there are also incentives for auditors to report more accurate hours. These incentives are necessary to produce an interior optimum, or an equilibrium where auditors do not report zero hours. That is, these countervailing pressures constrain auditors'

underreporting may not occur. Consequently, it is difficult to determine which audit engagements do and do not have underreporting. Thus, solutions such as adding a blanket amount to all budgets would result in over-charging clients on budgets that are appropriate.

[2] Another type of misreporting that could exist in this setting is overreporting or reporting more hours than one actually worked. Although overreporting incentives may be present, we know from prior literature that nearly two-thirds of auditors underreport (e.g., Lightner 1982). As such, although overreporting is an interesting avenue for future research, this study aims address the larger phenomenon of underreporting with the purpose of investigating a reporting procedure that reduces underreporting.

underreporting, such that they report a number of hours worked greater than zero. One incentive to report more accurately is that auditors want to be honest. The more auditors underreport, the more dishonest they feel. Another accuracy incentive is that an auditor may want to avoid causing tight budgets for subsequent audits because it will negatively impact future auditors. Finally, an additional accuracy incentive is the desire for reported information to appear realistic. Although managers cannot keep track of their staff's time on a fine scale (i.e., to the hour), they may have a rough idea of what reported hours should be. That is, managers may know when there are very large inaccuracies, which acts as a pressure for auditors to not underreport in excess.

The Theory of Mental Accounting

I turn to the theory of mental accounting to help motivate my hypotheses. Part of the theory of mental accounting extends prospect theory to incorporate multiple outcomes (i.e., multiple gains and/or multiple losses).[3] Specifically, in his theory of mental accounting, Thaler (1985) uses the value function from prospect theory to make predictions about individuals' preferences for experiencing aggregated or disaggregated gains and/or losses. To make these predictions Thaler (1985) determines the different amounts of utility one gains or loses when multiple positive outcomes and/or multiple negative outcomes are aggregated versus

[3] In 1979, Kahneman and Tversky published their paper outlining prospect theory. Prospect theory arguably presented a more descriptive theory of human behavior than the dominant theory of the time, expected utility theory. In creating prospect theory, Kahneman and Tversky (1979) replaced the utility function with a value function, which has three important features. First, the value function is based on changes, while the utility function is based on levels. That is, the value function predicts the change in utility individuals feel when they experience a change for the better (i.e., a gain) versus a change for the worse (i.e., a loss) with respect to some reference point. Second, the value function is concave in the domain of gains and convex in the domain of losses. That is, the difference between a $5 gain (loss) and a $10 gain (loss) seems greater than the difference between a $105 gain (loss) and a $110 gain (loss). Third, the loss function is steeper than the gain function. These three properties create the value functions used in Figure 1.

disaggregated.[4] Two implications from mental accounting are (1) individuals lose more utility when losses are disaggregated (delivered separately) than when losses are aggregated (delivered as one large loss), and (2) individuals gain more utility when gains are disaggregated (delivered separately) than when gains are aggregated (delivered as one large gain). In other words, individuals have different preferences when it comes to multiple gains versus multiple losses in that they prefer to aggregate multiple losses and disaggregate multiple gains.

Individuals prefer to aggregate losses because disaggregated losses are felt separately. That is, each separate loss begins at a new reference point on a new value function, which causes an individual to feel disutility from the steepest part of the value function multiple successive times (See Figure 1, Panel A). On the other hand, when losses are aggregated, they are felt as one large loss, and each loss within the aggregate simply shifts the impact over on the *same* value function. Consequently, when losses are aggregated, individuals feel the disutility from the flatter sections of the value function, which leads to a lower overall loss of utility for aggregated losses compared to disaggregated losses (See Figure 1, Panel B); thus, individuals prefer to aggregate losses.

Incorporating Reporting into the Theory of Mental Accounting

Reporting outcomes is integral to the field of accounting; however, in accounting there is often an intermediary between the actual outcome and the reported outcome (i.e., the reporter). If reporters experience gains or losses based on reported outcomes, and these reporters also have discretion or control over the reported outcome, then differences between the actual and reported information may arise. Mental accounting does not make predictions based on the distinction between actual and reported outcomes. That is, the current predictions of mental accounting

[4] In his 1985 paper, Thaler uses the terms integrate and segregate instead of aggregate and disaggregate. The terms are interchangeable. I choose to use the terms aggregate and disaggregate because their use is more common in the accounting literature to convey the construct of combining and separating items.

apply only once the outcomes are set and individuals are experiencing the outcomes. Mental accounting does not incorporate how individuals get to their reported outcomes or how aggregation and disaggregation might influence the actions individuals take to reach their desired outcomes. In this paper I aim to incorporate reporting into the theory of mental accounting to make predictions about when more or less misreporting will occur between aggregated and disaggregated reporting.

In the accounting literature, a number of studies examine how aggregated and disaggregated reporting influences individuals' preferences and behaviors. For example, Bonner et al. (2014), use mental accounting to examine manager preferences for aggregation versus disaggregation of income statement line items. My study extends Bonner et al. (2014) by developing predictions about whether individuals will take different actions when they report outcomes in an aggregated versus disaggregated format. My study also contributes to the literature on aggregated versus disaggregated analyst reporting, adding to papers such as Fennema and Koonce (2010), which uses mental accounting to predict firm and analyst decisions to aggregate versus disaggregate information in financial reports and voluntary disclosures. My study extends the analyst literature by predicting that misreporting incentives may result in differences between disaggregated and aggregated analyst reporting.

The most closely related paper to my study is Chen et al. (2015). Chen et al. (2015) investigates disaggregation and incentives in the management forecast setting. Because ambiguity exists in issuing management forecasts, if managers have incentives for higher earnings, they engage in motivated reasoning and unintentionally bias earnings forecasts. Chen et al. (2015) documents that as mangers issue more forecasts (i.e., the forecast is more disaggregated), there is more ambiguity, and therefore there is more unintentional bias in the

forecasts. My study is distinct from Chen et al. (2015) because I investigate a setting where individuals know a true outcome, so there is no ambiguity in reporting. When bias exists in my study it is due to intentional bias. Consequently, the theory in this paper generalizes to settings where individuals report outcomes, such as reporting components of net income or reporting information about a company's segments. These settings are distinct from the forecasting settings to which Chen et al. (2015) generalize.

An individual who faces misreporting incentives can avoid utility loss by misreporting. This utility loss occurs because an individual will experience disutility if he reports the true outcome as he loses out on incentives; thus by misreporting the individual avoids some of his utility loss. Turning to the value function, individuals who face misreporting incentives will misreport to shift their utility up on the value function. Due to the shape of the value function, misreporting or a horizontal shift right can result in different utility gains or mitigation of utility losses for the same amount of misreporting depending on how steep the value function is around the true outcome. Thus, because for multiple losses aggregated outcomes fall on the flatter sections of the value function while disaggregated outcomes fall on the steeper sections of the value function, individuals can avoid more utility loss by misreporting when reporting disaggregated versus aggregated information. If individuals avoid more utility loss from misreporting in disaggregated versus aggregated formats, then I predict misreporting will be greater when reporting is disaggregated. Because this prediction operates through misreporting incentives and mitigation of utility losses, if an individual does not face misreporting incentives, then the level of misreporting should not depend on the level of aggregation in reporting.

This theoretical extension can be applied to the audit hour setting as auditors face many incentives to manipulate their reported time and can be asked to report time in either an

aggregated or a disaggregated format. For example, a manager may ask for detailed hour data for each audit task or may ask for a more aggregated number such as the total number of hours spent working on the client. The incentives auditors face stem from the difference between the desired outcome (meeting the budgeted number of hours in this setting) and the true outcome (the actual number of hours worked). Auditors may foresee a loss if they do not obtain their desired outcome because they will lose out on the benefits stemming from the desired outcome. Thus, an auditor will underreport in order to report a smaller deviation from the desired outcome. When reporting disaggregated time compared to a disaggregated budget (hereafter "disaggregated format"), an auditor will foresee multiple losses because he compares the desired outcome to the true outcome multiple times. On the other hand when reporting aggregated time compared to an aggregated budget (hereafter "aggregated format"), an auditor will only foresee a single loss because he compares the desired outcome to the true outcome only once.

Auditors have the ability to manipulate their outcome by underreporting, thereby mitigating utility loss. Given the predictions developed above, if auditors face incentives to underreport and they report their time in a disaggregated format, I expect these auditors to underreport significantly more than auditors who report their time in an aggregated format or auditors who do not face underreporting incentives. I formalize these predictions in the hypothesis below:

> H1: Individuals will underreport more when they face underreporting incentives and report disaggregated time than when individuals face underreporting incentives and report aggregated time or when individuals do not face underreporting incentives.

Although I develop arguments to support my hypothesis, it is not a foregone conclusion that my predictions will apply in this setting. Auditors may simply underreport in equal amounts because the degree that individuals exceed the budget is the same regardless of whether hours are

aggregated or disaggregated. Further, gains and losses are often more directly associated with a monetary value. Reporting fewer hours only indirectly impacts pay via promotion opportunities (Agoglia et al. 2015). Consequently, to the extent that individuals do not equate manipulating one's outcome to gaining something of value, the effects predicted in this hypothesis may not exist.

Examining the Process of Aggregation in Reporting

As discussed above, I build on mental accounting by examining situations where individuals can manipulate their outcomes. In the audit hour setting, individuals can manipulate their outcomes by underreporting. I predict that the process through which individuals underreport is by mitigating utility loss. This process occurs because after underreporting, individuals should experience an outcome that results in higher utility than the outcome individuals would have experienced had they reported their true number of hours. In other words, there is utility loss individuals mitigate by underreporting. Thus, if my predictions building on mental accounting are descriptive, the utility loss that individuals mitigate by underreporting will act as a mediator in my analysis. I formalize this prediction in the below hypothesis:

H2: The utility loss individuals mitigate by underreporting mediates the relation between the level of aggregation, underreporting incentives, and the number of hours individuals choose to underreport.

The Cost of Aggregation

Although I propose that aggregated reporting reduces underreporting, the cost of reporting aggregated hours is reduced richness in audit hour data. For example, there is more detail or data richness in audit hour data if an auditor separately reports the time spent on multiple areas of an audit than if he reports the time as one aggregated number.[5] Rich and

[5] It is important to note that data richness is reduced only for some forms of aggregation. That is, although aggregating over audit tasks reduces data richness, aggregating over time may not reduce data richness. For

detailed data aids in decision-making; however, such data is most useful when it is accurate. Therefore, to address the trade-off between data richness and accuracy, I investigate another reporting procedure firms can use to gather rich data, while still maintaining the benefits of aggregation. Specifically, I investigate whether asking auditors to provide disaggregated percentages after reporting aggregated hours impacts the accuracy of aggregation.

Previously, I argued that auditors underreport because they foresee losses when they compare their desired outcome (i.e., the budget) to the true outcome and that by underreporting auditors mitigate utility losses that would otherwise experience had they reported their true time. It is possible that individuals may be able to distance themselves from the foreseen losses if they report disaggregated percentages. Specifically if instead of reporting disaggregated hours, individuals report aggregated hours followed by the percentage of time spent on the disaggregated areas, then individuals may not view the disaggregated percentages as multiple losses. This muted response to percentages is because individuals are not able to directly compare their hours to the budget, and so report more accurate hours as they cannot foresee losses. On the other hand, if individuals still view reporting percentages as reporting disaggregated losses, then individuals may underreport more when they report disaggregated percentages than when they do not report disaggregated percentages. Consequently, it is not clear whether reporting percentages is a reporting procedure that audit firms can use to obtain more detailed data while still maintaining the accuracy benefit of aggregated reporting. Thus, I present my examination as a research question that investigates the impact of percentage reporting on underreporting:

---

example, instead of reporting the time spent working on each audit task by day auditors can report aggregated time spent working on each audit task by week or month. This form of aggregation could still lead to more accuracy in reporting without loss of data richness.

RQ: Is the degree of underreporting equivalent between aggregated reporting with and without percentage reporting?

III. EXPERIMENT 1 – DESIGN, METHOD, AND RESULTS

Design and Participants

My first experiment investigates my first and second hypotheses. I collect data using a 2 × 2 (aggregated vs. disaggregated reporting × underreporting incentives present vs. absent) between-participants experiment. Participants begin by learning that they worked on eight tasks for a client, and that they must self-report their hours worked to their manager.[6] I manipulate whether participants feel incentives to underreport by informing participants that meeting budgets either is or is not important for the performance evaluation process and obtaining promotion opportunities. I manipulate aggregated or disaggregated reporting by providing participants with either (1) a total budgeted hours number and requiring participants to report one number for their total time spent working on the eight tasks, or (2) a budgeted number for each of the eight separate tasks and requiring participants to report eight separate numbers for their time spent working on each separate task.

Two-hundred and eighty workers from Amazon.com's Mechanical Turk (MTurk), an online work marketplace platform, participated in my experiment. Although I motivate my study using the audit setting, the task itself is not audit-specific. Libby, Bloomfield, and Nelson (2002) emphasize the need to recruit only the level of participant necessary to sufficiently test one's research question. Consequently, MTurk workers make appropriate participants for my

[6] I could have had participants perform an actual task as opposed to telling participants the amount of time they took to complete each task, however having participants perform an actual task would result in a confound. Specifically, aggregated and disaggregated budgets could influence performance in addition to reporting. That is, the pressure of being near the budget multiple times on each task versus being near the budget only once in total could impact how individuals perform on the tasks. My study investigates reporting alone, as such I wanted to hold constant performance, which I am able to do by telling participants exactly how much time they spent on each task.

experimental task because my study does not require any audit specific experience to complete.[7] Additionally, MTurk workers are a superior participant pool compared to auditors because I manipulate underreporting incentives as present and absent. In order to test the theory posed in this paper, I remove underreporting incentives to demonstrate that without underreporting incentives the effect of aggregation lessens. Given underreporting is a widespread phenomenon within the audit profession, auditors would not make appropriate participants because they would bring into the experiment their prior experience with underreporting incentives. With auditors, it would not be possible to examine a setting where underreporting incentives are absent. Therefore, I chose to use Mturk as participants for this study. I paid each MTurk participant $0.50 in exchange for completing the experiment.

Materials, Manipulations and Dependent Measure

The experiment was performed using Qualtrics, an online survey creator. All participants begin the study by learning that they worked on eight areas of a client during the past two weeks. Participants learn that they will submit their hours to the manager of the job, but that because they are salaried employees their submission will have no direct impact on their current pay.[8] I manipulate the presence versus absence of underreporting incentives by manipulating whether or not meeting the budget is important for performance evaluations and promotion opportunities.[9] That is, I tell participants in the underreporting incentives present conditions, "At your firm,

[7] On average, for Experiment 1, 37 percent of the participants had prior experience reporting hours compared to a budget. Importantly, the number of hours underreported did not vary by the level of familiarity with reporting hours ($F = 0.17$; $p = 0.92$ untabulated), and results are robust to including prior experience reporting hours as a covariate in all tests.

[8] I chose to inform participants that their pay is not impacted to mimic the Big 4 accounting firms, where reporting more hours does not increase pay. This salary component facilitates the widespread underreporting because there is a lack of significant benefits for reporting more hours (Sweeney and Pierce 2006).

[9] Prior research extensively documents the presence of underreporting and numerous causes such as wanting to look more efficient than one's peers, explicit pressure from managers or client pressures. For my main experiment I chose only two incentives for experimental participants to internalize (i.e., higher performance evaluations and more promotion opportunities). In the natural audit setting, auditors experience incentives beyond the two I use in my study. Thus, I suspect the hypothesized effects might be even stronger in the audit setting where participants have more underreporting incentives.

meeting budgets is critical in the performance evaluation process, and may impact your ability to be promoted in the future," and I tell participants in the underreporting incentives absent conditions, "At your firm, meeting budgets is not important in the performance evaluation process, and does not impact your ability to be promoted in the future."

Next all participants learn that they worked 97 total hours on eight separate tasks. Participants also learn the number of hours that they spent working on each individual task (i.e., 17 hours on Task 1, 12 hours on Task 2, 15 hours on Task 3, etc.). Below the information about the number of hours worked is a form that provides participants with both the number of hours that were budgeted for the tasks and a space for reporting their hours. The format of this form represents my second manipulation. Specifically, participants in the aggregated conditions view a single budgeted number and a form that requires participants to submit a single total number of hours (i.e., the total number of hours participants spent working on the client), while participants in the disaggregated conditions view eight separate budgeted numbers and a form that requires participants to submit a separate number of hours for each of the eight tasks (i.e., the number of hours spent working on Task 1, the number of hours spent working on Task 2, etc.).

My primary dependent variable is the number of hours that participants underreport. I calculate underreporting as the difference between the true total number of hours worked (97 hours) and the total number of hours that a participant reports. For the Aggregated conditions, the dependent variable is calculated by taking the difference between 97 hours and the single number reported. For the Disaggregated conditions, the dependent variable is calculated by taking the difference between 97 hours and the sum of the eight reported numbers.

To measure my mediator, *utility loss avoided*, I use a post experimental question that asks participants to what extent they agree or disagree with the statement, "If my manager was able to

view my true hours as opposed to my reported hours I would be unhappy." Participants respond using a 101-point scale with endpoints labeled *Strongly Disagree* (0) and *Strongly Agree* (100).[10] This question is meant to measure the utility an individual avoided losing by underreporting; the more individuals agreed with the statement the more utility the individual gained by underreporting. After providing an answer to the mediator measure, participants answer three questions: one attention check question; and two questions to determine whether the incentive manipulation was successful. At the end of the study participants answer demographic questions.

Results for Experiment 1

Manipulation and Attention Checks

To assess whether individuals attend to the underreporting incentives manipulation (i.e., whether or not they will obtain higher performance evaluations and promotion opportunities by underreporting), I ask participants two questions, which measure their perceptions of the benefits of reporting. On a 101-point scale with 0 being *Strongly Disagree* and 100 being *Strongly Agree*, participants disclose to what extent they agree that (1) they will be better off after reporting their time, and (2) they will receive positive benefits after reporting.[11] I perform a factor analysis on the two questions, and find that both questions load significantly on only one factor. Using this

[10] The construct I aim to measure is the amount of utility or happiness that participants gain by underreporting. In economics, the term happiness is sometimes used as a synonym for utility (see Clark, Frijters, and Shields (2008) for a discussion). Thus, I try to capture the change in utility between what participants could have hypothetically felt had they reported the true number of hours and what participants actually felt after reporting. When answering the post experimental question, participants have already reported their hours, and presumably have already gained utility or happiness by underreporting. Thus, I ask participants to think hypothetically about the happiness that would be lost had they instead reported their true hours (i.e., how unhappy they would be if they reported the true number of hours). Alternatively, I could have asked participants about their happiness before and after reporting hours, however this would have heightened participants' awareness of their utility prior to reporting, and could have resulted in demand effects. Thus I make the design choice to ask participants to compare their current state to a hypothetical state.

[11] The construct I aim to manipulate is whether individuals felt incentives or pressure to underreport. I chose not to ask about pressure because, to control for client and engagement characteristics, all participants were over the budget by the same amount. Individuals who are over the budget might all inherently feel pressure to underreport, and a question about pressure might not powerfully measure differences in my manipulated construct. Thus, to evaluate whether participants internalized the precise manipulation of obtaining promotion opportunities and better performance evaluations, I ask participants' about their feelings regarding benefits obtained after reporting.

factor I find that participants in the incentives present conditions agreed more with the statements that they would obtain benefits from reporting than participants in the incentives absent conditions (t = 1.52; p = 0.06, one-tailed, untabulated). Thus the manipulation of incentives appears to be successful.

I also ask participants an attention check question to ensure I only include participants who paid attention to the study. Specifically I tell participants, "This is an attention check question. Move the slider to Strongly Agree on the scale below". Participants must move a slider (i.e., a marker for their response) on a 101-point scale with endpoints *Strongly Disagree* (0) to *Strongly Agree* (100). Eight participants did not select *Strongly Agree* for this question. As such, I remove these participants from my analyses, which results in a final sample of 272 participants.[12]

Tests of H1

My first hypothesis predicts an ordinal interaction. Consequently, I use a single planned contrast to test whether the number of hours participants underreport fall into the pattern predicted by H1. Consistent with Buckless and Ravenscroft (1990), prior to performing my analysis, I select contrast weights for the purpose of testing a specific predicted pattern of results. I chose the following contrast weights for each experimental condition: -1 Incentives Present / Aggregated Format, +3 Incentives Present / Disaggregated Format, -1 Incentives Absent / Aggregated Format, and -1 Incentives Absent / Disaggregated Format. These contrast weights reflect my first hypothesis in the following ways. Specifically, the weights test whether individuals underreport more in the condition where they face underreporting incentives and report disaggregated time compared to the other three conditions where participants do not face

[12] As evidence that these individuals were not paying attention to the experimental materials, the average number of hours underreported by participants who failed the attention check was 42 hours. This average is over four times the average of all other participants.

underreporting incentives and/or report their time in an aggregated format. Also these weights test whether the effect of aggregation on the degree of underreporting is greater when incentives are present compared to when incentives are absent.

Table 1 Panel A reports the descriptive statistics for the number of hours underreported by experimental condition, and Figure 2 plots these descriptives. Table 1 Panel B reports the results of my contrast coding analysis. I find that the participants' average number of underreported hours falls in the hypothesized pattern supporting H1 (F = 6.75; p = 0.01). A semi-omnibus test confirms that the residual variance attributable to the main and interactive effects of aggregation and incentives after accounting for my planned contrast is not significant (F = 0.22; p = 0.80, untabulated). These results illustrate that auditors underreport more when they have incentives to underreport and report disaggregated time compared to when they have incentives to underreport and report aggregated time or when they do not have incentives to underreport. Thus, because auditors underreport less when they report aggregated time compared to disaggregated time, aggregation appears to reduce underreporting. These results also demonstrate that when auditors do not have incentives to underreport, aggregation does not significantly influence reporting, which suggests aggregation increases accuracy only when there are misreporting incentives.

Although my contrast coding analysis above supports H1, I also use simple effects to further investigate my findings. Table 1, Panel C reports the follow-up simple effects I use to validate my hypothesis. I perform three comparisons using the individuals who had underreporting incentives and were asked to report disaggregated time. These individuals underreported 11.99 hours on average, which is significantly greater than (1) individuals who had underreporting incentives and were asked to report aggregated time (11.99 vs. 6.45 hours; p

= 0.01 one-tailed), (2) individuals who did not have underreporting incentives and were asked to report disaggregated time (11.99 vs. 8.84 hours; p = 0.06 one-tailed), and (3) individuals who did not have underreporting incentives and were asked to report aggregated time (11.99 vs. 7.33 hours; p = 0.01 one-tailed). These simple effects provide support for H1. The most important simple effect comparison for the audit hour setting examines whether aggregation leads to less underreporting when incentives are present. Aggregation appears to help reduce underreporting as the difference between the two Incentives Present conditions is significant (11.99 vs. 6.45 hours; p = 0.01 one-tailed).

Test of H2

In support of my first hypothesis, I find that individuals underreport most when they have underreporting incentives and report their time in a disaggregated format. Building on mental accounting, I expect this relationship is driven by the utility individuals gain by underreporting. Thus my second hypothesis predicts that the utility that individuals gain by underreporting mediates the relationship between my independent variables, *level of aggregation* and *incentives to underreport*, and my dependent variable, *number of hours underreported*. Because underreporting is highest when individuals have underreporting incentives and report disaggregated time, I also expect the amount of utility individuals avoid losing by underreporting to be highest for these same individuals, compared to those reporting aggregated time and to those having no underreporting incentives. That is, consistent with H1, I expect an ordinal interaction for the mediating variable, *utility loss avoided*, with the following coefficients for each condition: -1 Incentives Present / Aggregated Format, +3 Incentives Present / Disaggregated Format, -1 Incentives Absent / Aggregated Format, and -1 Incentives Absent / Disaggregated Format.

Table 2 Panel A reports the descriptive statistics of my mediator, *utility loss avoided*, and Figure 3 presents my mediation analysis. According to Baron and Kenny (1986), to establish full mediation, the following three conditions must be met: (1) the manipulation of my independent variables, *level of aggregation* and *incentives to underreport*, significantly accounts for variations in the mediator, *utility loss avoided*, (2) variations in the mediator, *utility loss avoided*, significantly account for variations in the dependent variable, *number of hours underreported*, (3) when the mediator, *utility loss avoided*, is added as a control, the previously significant relationship between my independent variables, *level of aggregation* and *incentives to underreport*, and the dependent variable, *number of hours underreported*, is no longer significant.

As demonstrated in Figure 3, all conditions for mediation are present. The first condition of mediation is met as the manipulation of my independent variables, *level of aggregation* and *incentives to underreport*, significantly explain the variation in my mediating variable, *utility loss avoided* ($F = 10.52$; $p < 0.01$). The second condition of mediation is met as variations in *utility loss avoided* significantly account for variations in *number of hours underreported* ($F = 36.64$; $p < 0.01$). Finally, the third condition of mediation is met as the originally significant *level of aggregation* × *incentives to underreport* ordinal interaction ($F = 6.75$; $p = 0.01$) is no longer significant once *utility loss avoided* is added as a control ($F = 2.51$; $p = 0.11$). Consequently, it appears that the utility avoided by individuals by underreporting mediates relation between the level of aggregation, underreporting incentives, and the number of hours underreported, which is consistent with H2. Together, my results suggest that my predictions that build on mental accounting are descriptive in that (1) individuals who feel misreporting incentives misreport to

gain utility, and (2) when misreporting incentives exist, misreporting is greatest when reporting is disaggregated versus aggregated.

IV. EXPERIMENT 2 – DESIGN, METHOD AND RESULTS

Design and Participants

My second experiment investigates the research question of whether audit firms can maintain the accuracy benefits of aggregation, while also gaining some data richness through percentage reporting. Specifically, I examine whether the degree of underreporting will be impacted by percentage reporting by comparing the degree of underreporting for aggregated reporting with and without percentage reporting.

To investigate whether audit firms can utilize percentage reporting to maintain data richness, I collect data using a 2 × 2 (underreporting incentives present vs. absent × percentage reporting vs. no percentage reporting) between-participants experiment. Because I examine the effect of percentage reporting on the accuracy of aggregated reporting, in this experiment, all participants report aggregated time. As with Experiment 1, I manipulate whether or not participants feel incentives to underreport by telling half of the participants that meeting the budget is important in the performance evaluation process and for obtaining promotion opportunities, whereas the other half learns meeting the budget is not important in the performance evaluation process or for obtaining promotion opportunities. Also, to investigate if percentage reporting influences the accuracy benefits of aggregation, I manipulate whether or not participants report disaggregated percentages after reporting an aggregated number of hours.

Two-hundred and ninety-one workers from Mechanical Turk (MTurk), Amazon.com's online work marketplace, participated in my experiment. Again, I use MTurk workers for the same reasons outlined in Experiment 1. In exchange for completing the experiment, I paid each

MTurk participant $0.50. As with Experiment 1, I also ask participants an attention check question to ensure I only include participants who paid attention to the study. I use the same attention check question from Experiment 1, and I find nine participants fail the attention check. As such, I remove these participants from my analyses, which results in a final sample of 282 participants.

Materials, Manipulations and Dependent Measure

Similar to the previous experiment, participants performed this experiment using Qualtrics. Participants in this experiment view similar information to those in the prior experiment in the Incentives Present / Aggregated Format and Incentives Absent / Aggregated Format conditions; however, in this experiment half of the participants have an additional step after reporting their aggregated time where they report the percentage of time they spent working on each of the eight tasks. Additionally, the initial instructions for the participants in the percentage reporting conditions also include the additional phrase telling them they will also be reporting a percentage of time spent working on each task. When participants report percentages they are able to view their true number of hours as well as their reported number of hours.

My main dependent variable in Experiment 2 is the same as Experiment 1 – *number of hours underreported*. I calculate underreporting as the difference between the true total number of hours worked (97 hours) and the total number of hours participants report. This number represents how much participants bias their reported number of hours.

Results for Experiment 2

In my second experiment I examine whether there will be a difference in the degree of underreporting between individuals who do and do not report percentages. I find evidence that even though individuals are effectively reporting disaggregated hours by reporting disaggregated

percentages, reporting percentages does not significantly reduce the accuracy benefits of aggregated reporting as there is no main effect of percentage reporting or interaction between percentage reporting and incentives to underreport (both $F < 0.07$; both $p > 0.80$). Thus, it appears that firms can gain data richness through percentage reporting, without losing the accuracy benefits of aggregated reporting.

Follow up simple effects also support these results, as there is no significant difference between the conditions that did and did not report percentages for individuals with and without underreporting incentives (both $F < 0.07$; $p > 0.80$). Additionally, consistent with the first experiment, when participants were asked to report aggregated time, there is no significant difference in number of hours participants underreport between those who did and did not face underreporting incentives (without percentage reporting: $F = 0.22$; $p = 0.64$; with percentage reporting: $F = 0.35$; $p = 0.56$). Together these results suggest that, although some data richness is lost with aggregation, firms can gain back some detail with percentage reporting while maintaining the accuracy benefits of aggregated reporting.

V. CONCLUSION AND LIMITATIONS

The purpose of this paper is threefold. First, I propose a reporting procedure firms can use to reduce underreporting. In doing so, I build on the theory of mental accounting to demonstrate that when auditors feel incentives to underreport, aggregated reporting of audit hours will reduce the amount of underreporting (or result in less misreporting) compared to disaggregated reporting. Further, I document that when individuals do not feel underreporting incentives, aggregation does not significantly impact reporting. Second, I investigate the process of aggregation. My predictions build on mental accounting by incorporating an individual's ability to manipulate his outcomes. I test and find evidence consistent with my predictions by

performing a mediation analysis using the utility individuals avoid losing by underreporting as a mediator for the relationship between aggregation, underreporting incentives and number of hours underreported. These results could be applied more broadly to other reporting settings as this study suggests that disaggregation results in systematic misreporting when there are incentives to misreport. Third, not only do I propose aggregation as a tool for improved audit hour accuracy, I also propose disaggregated percentage reporting as a reporting procedure that allows audit firms to capture some detailed data, while still maintaining the accuracy benefits of aggregation.

My results are important for auditors and the PCAOB. Audit firms and teams use reported hours to assess auditors at all levels. Aggregating audit hours could lead to more accurate and fair assessments of auditors. Accurate audit hours could also lead to more accurate budgets. If budgets are appropriate, audit quality will be higher as auditors will be less likely to take audit quality reducing actions that tight budgets cause such as performing superficial reviews, prematurely signing off on workpapers, and accepting weak client explanations. Additionally, because audit firms currently use audit hours in their audit quality measures and the PCAOB is considering the use of audit hours in several of their audit quality indicators, this study suggests that disaggregated audit hour data should be used with caution because such disaggregation could lead to inaccurate assessments of audit quality.

I acknowledge that there are limitations to this study. One limitation of my study is the fact that I do not use auditors as participants. Although I argue that MTurk workers make appropriate participants for this non-audit task, it is possible that auditors respond differently to aggregation than non-auditors. Another limitation, faced by many experiments, is that my setting is stylized and abstracted from the real world. I consider this to also be a benefit of my study, as I

can hold constant factors that might impact auditor reporting; however, I only examine settings where auditors are over-budget on multiple areas. Mental accounting provides other predictions for various combinations of gains and losses (e.g., a large loss and a small gain or a small loss and a large gain). Thus, future research can investigate misreporting pressures and reporting bias resulting from the aggregation and disaggregation of other gain and loss combinations, such as settings where hours shifting may occur.

References

Agoglia, C. P., R. C. Hatfield, T. A. Lambert. 2015. Audit team time reporting: An agency theory perspective. *Accounting, Organizations and Society* 44: 1-14.

Baron, R. M., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology* 51(6): 1173–1182.

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.

Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast Coding: A Refinement of ANOVA in Behavioral Analysis. *The Accounting Review* 65(4): 933–945.

Buchheit, S., W. R. Jr. Pasewark, and J. R. Strawser. 2003. No need to compromise: Evidence of public accounting's changing culture regarding budgetary performance. *Journal of Business Ethics* 42: 151–163.

Chen, C. X., K. M. Rennekamp, and F. H. Zhou. 2015. The effects of forecast type and performance-based incentives on the quality of management forecasts. *Accounting, Organizations and Society* 46: 8-18.

Clark, A. E., P. Frijters, and M. A. Shields. 2008. Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature* 46(1): 95-144.

Coram P., J. Ng, and D. R. Woodliff. 2004. The Effect of Risk of Misstatement on the Propensity to Commit Reduced Quality Acts under Time Budget Pressure. *Auditing: A Journal of Practice & Theory* 23: 437-488.

Deloitte. 2016. Deloitte US 2016 Transparency Report. Available at: http://www2.deloitte.com/us/en/pages/regulatory/transparency-report.html.

Donnelly, D. P., Quirin, J. J., and O'Bryan, D. 2003. Auditor acceptance of dysfunctional audit behavior: An explanatory model using auditors' personal characteristics. *Behavioral Research in Accounting* 15: 87-110.

EY. 2015. Transparency Report 2015 EY US. Available at: http://www.ey.com/Publication/vwLUAssets/ey-us-transparency-report-2015/$FILE/ey-us-transparency-report-2015.pdf.

Fennema, M. G. and L. Koonce. 2010. Mental accounting in financial reporting and voluntary disclosure. *Journal of Accounting Literature* 29: 1-29.

House of Lords (HOL). 2010. Select committee on Economic Affairs. Auditors: Market concentration and their role, Volume II: Evidence. Published by the Authority of the House of Lords.

Kahneman, D. and A. Tversky. 1979. Prospect theory: An analysis of decision under uncertainty. *Econometrica* 47(2): 263-292.

Kelley, T. and L. Margheim. 1990. The Impact of Time Budget Pressure, Personality, and Leadership Variables on Dysfunctional Auditor Behavior. *Auditing: A Journal of Practice & Theory* 9(2): 21-42.

KPMG. 2015. Communicating Audit Quality KPMG International Transparency Report. Available at: https://assets.kpmg.com/content/dam/kpmg/pdf/2015/12/2015-international-transparency-report.pdf.

Libby, R., Bloomfield, R., and M. Nelson. 2002. Experimental Research in Financial Accounting. *Accounting, Organizations and Society*. 27(8): 775-810.

Lightner, S. M. 1982. Does Your Firm Have an Underreported Chargeable Time Problem? *The Practical Accountant* 15(4): 56-60.

Lightner, S. M., S. J. Adams, K. M. Lightner. 1982. The influence of Situational, Ethical and Expectancy Theory Variables on Accountants' Underreporting Behavior. *Auditing: A Journal of Practice & Theory* 2(1): 1-12.

Lightner, S. M., J. J. Leisenring, and A. J. Winters. 1983. Underreporting chargeable time. *Journal of Accountancy* 155: 52-57.

Margheim, L. and K. Pany. 1986. Quality Control, Premature Signoff and Underreporting of Time: Some Empirical Findings. *Auditing: A Journal of Practice & Theory* 5(2): 50-63.

McNair, C. J. 1991. Proper Compromises: The Management Control Dilemma in Public Accounting and its Impact on Auditor Behavior. *Accounting, Organizations and Society* 16(7): 635-653.

Nelson, M. W., and K. 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.

Otley, D. T., and B. J. Pierce. 1996. The Operation of Control Systems in Large Audit Firms. *Auditing: A Journal of Practice & Theory* 15(2): 65-84.

Ponemon, L. A. 1992. Auditor underreporting of time and moral reasoning: An experimental lab study. *Contemporary Accounting Research* Vol 9(1): 171-189.

PricewaterhouseCoopers (PwC). 2015. Our focus on audit quality. Available at: http://www.pwc.com/us/en/audit-assurance-services/publications/assets/pwc-2015-audit-quality-report.pdf.

Public Company Accounting Oversight Board (PCAOB). 2015. Concept Release on Audit Quality Indicators.

Public Oversight Board (POB). 2000. The panel on audit effectiveness report and recommendations. Stamford, CT.

Sweeney, B., and B. Pierce. 2006. Good hours, bad hours and auditors' defence mechanisms in audit firms. *Accounting, Auditing & Accountability Journal* 19(6): 858-892.

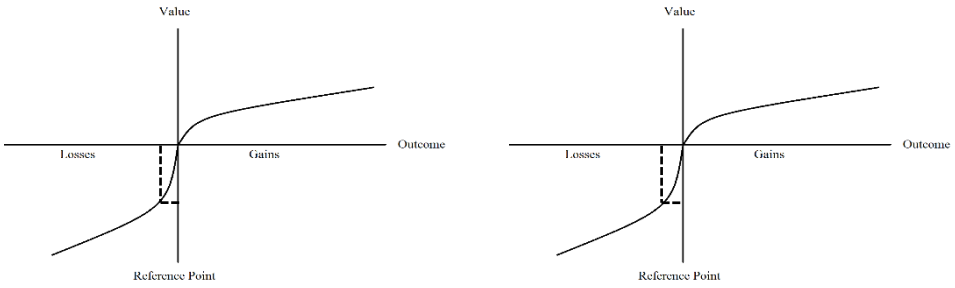
Taylor, E. Z., M. B. Curtis, and L. Chui. 2012. Staff Auditors' Observations of Questionable Peer Behavior: The view from the other side. *The CPA Journal* June: 66-71.

Thaler, R. H. 1985. Mental Accounting and Consumer Choice. *Marketing Science* 27(1): 15-25.

F I G U R E 1

Illustration of Disaggregation and Aggregation of Multiple Losses using the Value Function

Panel A: Two Equivalent Losses Disaggregated



Panel B: Two Equivalent Losses Aggregated

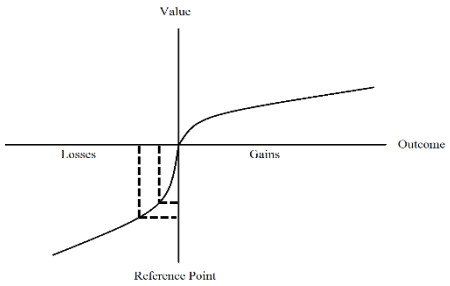


Figure 1 illustrates Thaler's (1985) implications of aggregating and disaggregating multiple losses within mental accounting.

Panel A represents two losses that are disaggregated (i.e., experienced as two separate losses). The panel displays two value functions because when losses are disaggregated they are felt separately and each loss is measured from a new reference point. The total value or utility lost is the sum of two values from each value function where the dotted line meets the vertical value axis.

Panel B represents two losses that are aggregated (i.e., experienced as one larger loss). The panel displays one value function because when losses are aggregated they are felt as one large loss. The losses are added together and measured from a single reference point. The total value or utility lost is the one point where the second dotted line meets the value axis. It is important to note that the second dotted line represents the additional utility or value lost when the second loss is aggregated. Even though the value of the second loss is the same, the amount of utility or value lost is much less than the first loss due to the convexity of the value function in the loss domain.

Together panels A and B demonstrate that the utility or value lost when losses are disaggregated is greater than the utility or value lost when losses are aggregated.

F I G U R E 2

Observed effects of Incentives and Aggregation on Participants' Number of Hours Underreported

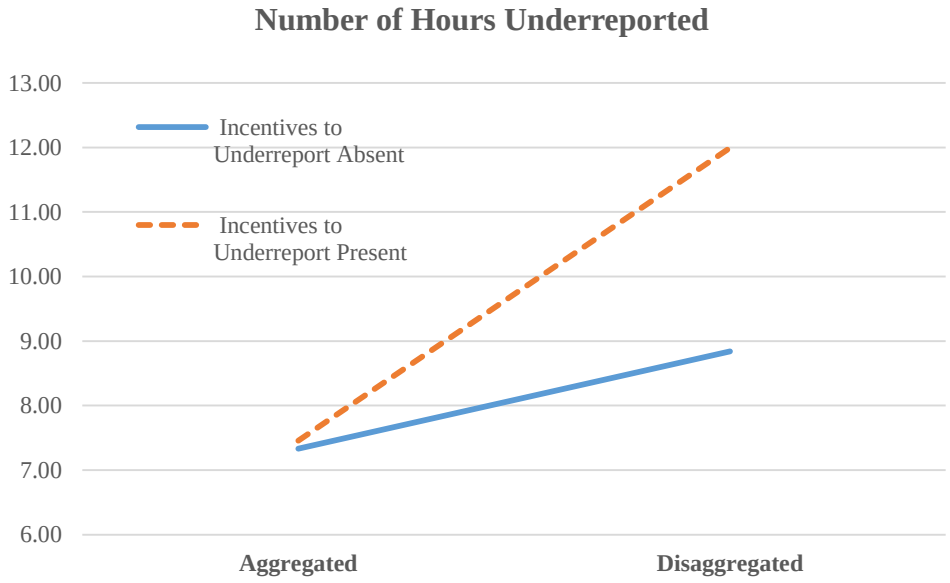


Figure 2 illustrates the means for my main dependent variable, *Number of Hours Underreported*, by experimental condition, as reported in Table 1 Panel A. In my experiment, I manipulate the incentives to underreport as present vs. absent and the level of aggregation in the reporting format participants used to disclose their hours to their manager as aggregated vs. disaggregated. These two manipulations result in four treatment conditions.

F I G U R E 3

Mediation Analysis

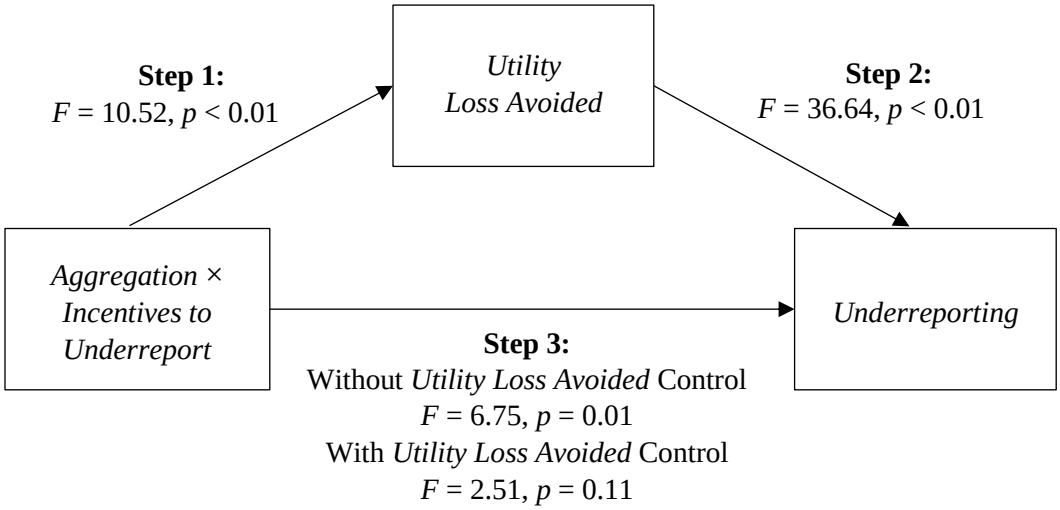


Figure 3 illustrates the mediation analysis discussed in Section III. This mediation analysis uses the data from Experiment 1. The mediator *Utility Loss Avoided* is individuals' responses on a 101-point Likert scale with endpoints *Strongly Disagree* (0) and *Strongly Agree* (100) denoting agreement with the statement, "If my manager was able to view my true hours as opposed to my reported hours I would be unhappy". All p-values in Figure 3 are one-tailed equivalents to represent directional predictions.

T A B L E 1

Experiment 1 - Number of Hours Underreported

Panel A: Descriptive Statistics for *Number of Hours Underreported* (mean, [standard deviation])

	Incentives to Underreport	
Level of Aggregation	*Absent*	*Present*
Aggregated	7.33 [14.743] *n* = 72	7.45 [8.181] *n* = 66
Disaggregated	8.84 [10.746] *n* = 62	11.99 [11.119] *n* = 72

Panel B: Contrast Coding with Dependent Variable - *Number of Hours Underreported*

Source of Variation	**SS**	**df**	**MS**	**F-statistic**	**p-value**
Model Contrast	896.67	1	896.67	6.75	0.01
Error	35604.00	268	132.85		

Panel C: Simple Effects with Dependent Variable - *Number of Hours Underreported*

Source of Variation	**df**	**F-statistic**	**p-value**
Effect of Aggregation given Underreporting Incentives Present	1	5.34	0.01†
Effect of Aggregation given Underreporting Incentives Absent	1	0.57	0.45
Effect of Incentives given Aggregated Reporting Format	1	0.00	0.95
Effect of Incentives given Disaggregated Reporting Format	1	2.50	0.06†
Disaggregated Reporting with Incentives Present compared to Aggregated Reporting with Incentives Absent	1	5.88	0.01†

Table 1 presents the results for my main dependent variable from Experiment 1 - *Number of Hours Underreported*. *Number of Hours Underreported* is calculated as the difference between the true total number of hours worked (i.e., the number provided to all participants in the materials - 97 hours), and the number that participants reported to their manager as having worked. For the aggregated conditions *Number of Hours Underreported* would be 97 hours minus the one number submitted to the manager, while for the disaggregated conditions *Number of Hours Underreported* would be 97 hours minus the sum of all eight numbers submitted to the manager.

Panel A presents descriptive statistics for *Number of Hours Underreported* by experimental condition. In my experiment, I manipulate the incentives to underreport by telling participants meeting the budget either is or is not important, and I manipulate the level of aggregation in the reporting format that participants must use to disclose their hours to the manager as either aggregated as one number or disaggregated as eight separate numbers. These two manipulations result in four treatment conditions.

Panel B presents the contrast coding for the planned contrasts. Contrast coefficients are +3 for the Incentives Present / Disaggregated Reporting condition, -1 for the Incentives Present / Aggregated Reporting condition, -1 for the Incentives Absent / Disaggregated Reporting condition, and -1 for the Incentives Absent / Aggregated Reporting condition. A semi-omnibus test confirms that the residual variance attributable to the main and interactive effects of aggregation and incentives after accounting for my planned contrast is not significant ($F = 0.22$; $p = 0.80$, untabulated).

Panel C presents the more granular follow-up simple effects tests used in examining my hypotheses. † p-values are one-tailed equivalents for directional predictions, and all other p-values are two-tailed equivalents.

T A B L E 2

Experiment 1 - Utility Loss Avoided

Panel A: Descriptive Statistics for *Utility Loss Avoided* (mean, [standard deviation])

	Incentives to Underreport	
Level of Aggregation	*Absent*	*Present*
Aggregated	26.90 [31.934] *n* = 72	35.02 [33.831] *n* = 66
Disaggregated	26.23 [30.354] *n* = 62	44.08 [30.269] *n* = 72

Panel B: Contrast Coding with Dependent Variable - *Utility Loss Avoided*

Source of Variation	**SS**	**df**	**MS**	**F-statistic**	**p-value**
Model Contrast	11431.92	1	11431.92	10.52	< 0.01
Error	291317.60	268	1087.01		

Panel C: Simple Effects with Dependent Variable - *Utility Loss Avoided*

Source of Variation	**df**	**F-statistic**	**p-value**
Effect of Aggregation given Underreporting Incentives Present	1	2.61	0.05†
Effect of Aggregation given Underreporting Incentives Absent	1	0.01	0.91
Effect of Incentives given Aggregated Reporting Format	1	2.09	0.15
Effect of Incentives given Disaggregated Reporting Format	1	9.77	< 0.01†
Disaggregated Reporting with Incentives Present compared to Aggregated Reporting with Incentives Absent	1	9.78	< 0.01†

Table 2 presents the results for my mediating variable from Experiment 1 - *Utility Loss Avoided*. *Utility Loss Avoided* is measured using the extent to which participants agree or disagree with the statement, "If my manager was able to view my true hours as opposed to my reported hours I would be unhappy." Participants respond using a 101-point scale with endpoints labeled *Strongly Disagree* (0) and *Strongly Agree* (100).

Panel A presents descriptive statistics for *Utility Loss Avoided* by experimental condition. In my experiment, I manipulate the incentives to underreport by telling participants meeting the budget either is or is not important, and I manipulate the level of aggregation in the reporting format that participants must use to disclose their hours to the manager as either aggregated as one number or disaggregated as eight separate numbers. These two manipulations result in four treatment conditions.

Panel B presents the contrast coding for the planned contrasts. Contrast coefficients are +3 for the Incentives Present / Disaggregated Reporting condition, -1 for the Incentives Present / Aggregated Reporting condition, -1 for the Incentives Absent / Disaggregated Reporting condition, and -1 for the Incentives Absent / Aggregated Reporting condition. A semi-omnibus test confirms that the residual variance attributable to the main and interactive effects of aggregation and incentives after accounting for my planned contrast is not significant ($F = 1.28$; $p = 0.28$, untabulated)

Panel C presents the more granular follow-up simple effects tests used in examining my hypotheses. † p-values are one-tailed equivalents for directional predictions, and all other p-values are two-tailed equivalents.

T A B L E 3

Experiment 2 - Number of Hours Underreported

Panel A: Descriptive Statistics for *Number of Hours Underreported* (mean, [standard deviation])

	Incentives to Underreport	
Percentage Reporting	*Absent*	*Present*
Absent	7.33 [11.231] *n* = 75	8.17 [11.500] *n* = 72
Present	6.87 [12.021] *n* = 67	7.97 [8.329] *n* = 68

Panel B: Analysis of Variance with Dependent Variable - *Number of Hours Underreported*

Source of Variation	**SS**	**df**	**MS**	**F-statistic**	**p-value**
Percentage Reporting	7.75	1	7.75	0.07	0.80
Incentives	66.08	1	66.08	0.56	0.46
Percentage Reporting × Incentives	1.30	1	1.30	0.01	0.92
Error	1.30	1	1.30		

Panel C: Simple Effects with Dependent Variable - *Number of Hours Underreported*

Source of Variation	**df**	**F-statistic**	**p-value**
Effect of Percentage Reporting given Underreporting Incentives Present	1	0.01	0.92
Effect of Percentage Reporting given Underreporting Incentives Absent	1	0.07	0.80
Effect of Incentives given Percentage Reporting Format	1	0.35	0.56
Effect of Incentives given No Percentage Reporting Format	1	0.22	0.64

Table 3 presents the results for my main dependent variable from Experiment 2 - *Number of Hours Underreported*. *Number of Hours Underreported* is calculated as the difference between the true total number of hours worked (i.e., the number provided to all participants in the materials - 97 hours), and the number that participants reported to their manager as having worked. All participants in this experiment reported their hours in an aggregated format, as such *Number of Hours Underreported* was calculated as 97 hours minus the one number submitted to the manager.

Panel A presents descriptive statistics for *Number of Hours Underreported* by experimental condition. In my experiment, I manipulate the incentives to underreport by telling participants meeting the budget either is or is not important, and I manipulate whether or not participants report percentages for the amount of time spent on each task after reporting their aggregated time. These two manipulations result in four treatment conditions.

Panel B presents the Analysis of Variance with the dependent variable being *Number of Hours Underreported*.

Panel C presents the more granular follow-up simple effects tests used in examining my hypotheses. All p-values are two-tailed equivalents.