The impact of procedural and distributive justice on satisfaction and manufacturing performance: a replication of Lindquist (1995) with a focus on the importance of common metrics in experimental design

Tim M. Lindquist¹ · Alexandra Rausch²

Accepted: 1 February 2021 / Published online: 20 March 2021
© The Author(s) 2021

Abstract

This paper replicates Lindquist's (Lindquist, *Journal of Management Accounting Research* 7:122–147, 1995) seminal research introducing the concepts of justice to the accounting literature. We use organizational justice theory, as did he, to replicate his study and, in doing so, question some findings of partial replications and extensions done over the past 25 years. We do this because work built off his study has challenged some of his findings. These challenges, we believe, have resulted from most researchers using different research metrics than did Lindquist. Many of these extensions have also used a mental-based task, instead of a manual-based one, in their experiments. We believe this constrains the ability to draw inferences and conclusions from this subsequent research. We further believe this constraint extends to much of the experimental research in the social sciences.

In our research we replicate exactly Lindquist's (Lindquist, *Journal of Management Accounting Research* 7:122–147, 1995) operationalizations of voice and vote and measure dependent outcomes for four of the same conditions he investigated. In contrast to most of other follow-up studies, we find, as did Lindquist, that having a voice only leads to significantly enhanced satisfaction with high-stretch targets, as compared to having a vote only or no input. We also corroborate Lindquist's (Lindquist, *Journal of Management Accounting Research* 7:122–147, 1995) result that having a voice only leads to significantly greater satisfaction with the experimental task, as compared to participants with a vote only or no input. Additionally, unlike Lindquist (Lindquist, *Journal of Management Accounting Research* 7:122–147, 1995), we find participants allowed only a voice significantly outperform participants with a vote only and no input. We thus support Lindquist's findings of a fair process effect for voice and perceptions of pseudo-participation related to vote.

Keywords Distributive justice · Procedural justice · Stretch targets · Experimental tasks · Replication study

JEL Classifications C1 Econometric and Statistical Methods and Methodology: General · C18 Methodological Issues: General · C9 Design of Experiments · C90 General · L6 Industry Studies: Manufacturing · L60 General

1 Introduction

Participative budgeting is one of the most examined topics in budget research (Dau-moser et al. 2018). A review of management accounting literature shows that organizational justice is among factors that have gained importance in recent years (Liessem et al. 2015). Lindquist (1995), who pioneered justice research in accounting (Indriani 2015), proposes that an antecedent to participative budgeting is a general desire to maintain justice in incentive-based compensation systems. He posits that employers seek increased employee participation in the development of incentive-based compensation to introduce fairness to the budgeting. He indicates that perceptions of justice involving not only the level of difficulty of performance standards (distributive justice/fair and unfair standards) but the process used to determine performance budgets (procedural justice/voice and vote) impact satisfaction and manufacturing performance. When individuals have a voice, they can participate by expressing their opinions and feelings. When they have a vote, they can endorse preferred standards (in budgeting). In his experimental study, Lindquist finds that voice, a form of low process control, enhances satisfaction with stretch targets and with the experimental task itself, while vote, a form of high procedural control, does not and is perceived as pseudo-participation when there are unattainable standards (i.e., standards higher than those for which the individual voted). No evidence is found for a significant performance effect. Since Lindquist employs a manual-based experimental task—building toy-castles out of plastic *Loc Blocs*®—he argues that his results apply to manufacturing budgeting.

Since publication, Lindquist (1995) has received over 220 citations, many of them extensions of his work, and yet only a few of these use an experimental design, as he did, to partially replicate and extend the work (Libby 1999; Libby 2001; Chow et al. 2001; Libby 2003; Byrne and Damon 2008; Francis-Gladney et al. 2008; Nahartyo 2013; Kelly et al. 2015; Gomez-Ruiz and Rodriguez-Rivero 2018). On top of this, the results of these experimental studies are sometimes inconsistent both with the findings of Lindquist (1995) and each other. For example, while some extensions have found similar impacts for voice (Byrne and Damon 2008; Chow et al. 2001; Libby 1999) as Lindquist (1995), others have found aspects of voice to negatively affect dependent measures (Nahartyo 2013). Also, Lindquist's (1995) failure to find a performance effect for any form of procedural justice has been challenged by Libby (1999), Chow et al. (2001) for Chinese nationals, Byrne and Damon (2008), and Kelly et al. (2015), who all found performance effects.

The purpose of a replication is to confirm findings. While many of these extensions explore additional antecedents of participative budgeting, they also refer to Lindquist (1995), noting how in their studies, voice or vote or both voice and vote affect satisfaction and performance differently. Fortunately, failure to replicate results of experimental studies provides an opportunity to explore the reasons for the different findings (Byrne and Damon 2008). In our study, we propose that these inconsistencies between Lindquist's original findings and subsequent ones stem from variations in research design. To support this proposition, our study partially replicates Lindquist (1995). We replicate some aspects of his study exactly, but we also introduce a methodologically consistent change (Darley 2000): our manual task requires participants to produce quilts out of paper scraps in our experimental design.

Replications enjoy a rich history in the natural sciences but are uncommon in the social sciences (Hensel 2019). Too often, in the social sciences, they are viewed as deficient in ingenuity. However, results should be replicated before extensions into new areas are proposed (Otley and Pollanen 2000). Without replications, social scientists can be prone to falsely infer findings from studies with limited sample sizes or errors in experimental design. This is undesirable, as findings from social science research are often employed to indicate preferable human behavior. The tide is turning, however, and replications are beginning to appear in psychology, economics and management (Block and Kuckertz 2018; Brandt et al. 2014; Motyl et al. 2017). We choose to replicate Lindquist (1995) because his seminal work encompasses the three dimensions of social science listed above—psychology, economics and management science.

Lindquist found a fair process, such that allowing participants to voice their feelings between production rounds (voice), enhances satisfaction with high stretch budgets more than higher levels of procedural justice (vote) or both (voice and vote). As mentioned above, other subsequent studies have been inconclusive in supporting this relationship. This stems from their operationalizations of voice and vote. Lindquist's (1995) vote allowed participants to choose a standard with the knowledge that it might not be accepted by management. Some follow-up studies unfortunately include a component of vote in what they term voice (Libby 1999; Bryne and Damon 2008). Also, while some research supports Lindquist's contention that workers may perceive vote as pseudo-participation (Francis-Gladney et al. 2008; Nahartyo 2013; Gomez-Ruiz and Rodriguez-Rivero 2018), other studies claim that voice leads to the same negative effect on satisfaction and performance (Libby 1999; Byrne and Damon 2008). Finally, Lindquist (1995) finds no performance effects for any measures of procedural justice. While some extensions concur (e.g., Kelly et al. 2015), Libby (1999) does find performance enhancements when voice is combined with an explanation for unattainable performance standards. Note again, though, that she operationalized voice differently than Lindquist (1995) and uses a *mental*-based experimental task.

We contend that the type of task used in prior extensions impacts their ability to challenge Lindquist's (1995) findings. Birnberg and Nath (1968), over 50 years ago, were the first to recognize a dichotomy between manual (physical) experimental tasks and mental ones. They stated clearly that a representative

task, simulating manufacturing, should recreate the routine nature of industrial settings. Shin and Grant (2019) support this contention, noting that a representative manufacturing task should induce fatigue and boredom. The problem is, except for a few studies (Chow et al. 2001; Lindquist 1995), all experimental research examining the impacts of justice/injustice on manufacturing workers has employed Chow's (1983) mental-based symbol decoding task. Mental-based tasks like this can generate a higher level of interest and intrinsic motivation than manual ones. And with them, effort may become its own reward (Amabile 1993; Keller and Bless 2008).

Our experimental findings mirror those of Lindquist (1995). We also find that allowing participants to help set their performance standards leads to positive outcomes at low process control levels (voice) but backfires at high levels (vote). Having a voice enhances budget and task satisfaction, as compared to having a vote. We also find, as did Lindquist (1995), that experimental participants perceive vote to be a countervailing form of pseudo-participation, which results in a decrease in satisfaction. Finally, while Lindquist (1995) could not find performance effects with the toy castle task, our paper quilt task shows a significant positive effect for performance, such that participants with a voice outperform those without one. Additionally, participants with only a voice outperform those with only a vote.

Our findings imply several things. First, we show that, if a study aims to replicate or extend previous work, adopting an appropriate methodology with common metrics is essential. We operationalize procedural justice—voice and vote—in the exact same manner as Lindquist (1995) and find the same relationships between it, at low (voice) and high (vote) levels of process control, and satisfaction, as did he. Second, we deliberately use a different manual task (paper quilt making) but keep the type of task in line with the recommendations of Birnberg and Nath (1968) and Shin and Grant (2019) that representative manufacturing tasks should be repetitive and mundane. That likely explains why our voice and vote lead to similar enhancements in satisfaction in our study, as compared to Lindquist (1995), and why extensions using symbol decoding have had varied findings. Third, in employing paper quilt making, we offer the management accounting literature a valid manual-based task to use in experimental research. Finally, our study, unlike Lindquist's (1995), finds performance enhancements for participants with voice. Individuals given a voice outperformed those with none, with vote only, or with no input, even when piece-rate standards assigned to them were unattainable. Because we matched all other parameters in our study to those of Lindquist (1995), this might suggest that making paper quilts is a superior task to building castles from plastic pieces. Quilt making is much simpler than castle building, and perhaps the impact of learning played a smaller role.

The remainder of the paper is organized as follows. Section 2 reviews the literature and develops the hypotheses. Section 3 describes the method used to conduct this replication and extension, and Sect. 4 presents results. Section 5 provides a discussion and conclusions and addresses limitations of the study as well as directions for future research.

2 Literature review and development of hypotheses

2.1 Replication of experimental studies in management (accounting) research

Replication studies in the natural sciences are plentiful and often repeat laboratory experiments in the exact same way as the original study (Hensel 2019). Replications are encouraged in medical and pharmacological fields to ensure findings' efficacy and safety. But what of the social sciences? For decades, a stigma has existed toward replications in the social sciences. Here, a consensus seems to be that publishable manuscripts must extend previous research and offer new perspectives. While there are benefits to this paradigm, it creates the danger of making assumptions based on studies with limited sample sizes or errors (Schmidt 2009). Fortunately, Block and Kuckertz (2018) found that replications of experimental studies are increasingly common in social sciences, such as psychology, economics and management. Brandt et al. (2014) recognize that, since 2012, prestigious psychology journals have been willing to publish both failed and successful replications. This is vital, as other psychologists have demonstrated that too many published findings are not robust (Ortmann 2017) and perform poorly when replicated (Open Science Collaboration 2015). In economics, scholars have only recently begun to address the lack of replications (Motyl et al. 2017). Encouragingly, replications have been published in reputable academic journals and show that experiments, particularly in behavioral economics, seem to fare relatively well when reproduced (Camerer et al. 2016). Finally, management science lacks replications of key research (Kepes et al. 2014; Makel and Plucker 2014), which is troubling because this line of research is grounded in human behavior (Morrison et al. 2010). While a few management journals (e.g., *Management Review Quarterly*) have recently begun to publish replications (Block and Kuckertz 2018), the number is still low (Hensel 2019). In management accounting, reanalysis of key works is rare, and only now are top journals calling for replication studies.

For our study, we choose to replicate Lindquist (1995), because the author's work encompasses all three social science paradigms discussed above: psychology, economics, and management science. It borrows constructs of procedural justice and referent cognitions studied in psychology (e.g. Lind et al. 1990; Skarlicki and Folger 1997; Tyler 1994; Tyler and Blader 2003) and implements incentives for a real effort task, as common in behavioral economics (e.g. Carpenter and Huet-Vaughn 2019). It further addresses performance in connection with budget setting, which is a prominent issue in both general management and management accounting (Daumoser et al. 2018; Derfuss 2016; Liessem et al. 2015). Further, Lindquist's (1995) seminal work, which pioneered justice research in accounting (Indriani 2015), is broadly cited. That may explain why a number of authors have already crafted experimental studies to try to extend it and investigate the impact of justice on varied dependent measures (Libby 1999; Libby 2001; Chow et al. 2001; Libby 2003; Byrne and Damon 2008; Francis-Gladney et al. 2008; Nahartyo 2013; Kelly, Webb and Vance 2015; Gomez-Ruiz

and Rodriguez-Rivero 2018). Also, numerous other authors have developed field studies and surveys to further investigate the impact of justice issues in accounting (Lau and Lim 2002a; Lau and Lim 2002b; Lau and Tan 2005; Lau and Tan 2006; Chong and Strauss 2017; Zainuddin and Isa 2019; Sudarwan 2019; Habran and Mouritsen, 2020; Safkaur and Pangayow 2020). Since our replication of Lindquist (1995) has an experimental design, we focus on the nine experimental extensions listed above in the remainder of this literature review.

2.2 The operationalization of procedural justice in experimental participative budgeting studies

Two key findings of Lindquist's original study arouse interest because they are not entirely supported in follow-up studies or significantly contradict theory. To begin, Lindquist finds that voice, as a form of low process control without decision control, enhances satisfaction in the presence of stretch targets. This increase in satisfaction is known as the "fair process effect" and is independent of the perception of budget attainability. Hence being allowed to express opinions, thoughts, and feelings boosts satisfaction, even with stretch targets. In his original study, Lindquist (1995) provides a basis for the operationalization of procedural justice as two independent variables (voice and vote) by employing a continuum of participation based on the Vroom-Yetton model.¹ This continuum ranges from no input (i.e., having neither voice nor vote) to voice indicating low process control, vote, and finally to voice and vote (i.e., the highest level of control). In Lindquist's study, voice refers to individuals' opportunities to express their opinions, thoughts, feelings, etc., and to comment on standard setting, while vote refers to individuals' opportunities to communicate their preferred standards to superiors. Secondly, Lindquist (1995) finds that, when participants who receive unattainable standards are given an opportunity to vote (a higher form of process control), they form referent cognitions of what their standard might have been (i.e., an attainable standard). These referent cognitions lead to perceptions of pseudo-participation, which result in declines in both budget and task satisfaction, where there should be increases from the higher form of process control (vote). As to Lindquist's (1995) investigation on the impact of voice as a form of low process control on performance, results show that enhancing procedural justice by allowing employees to voice their opinions, thoughts, feelings, etc., does not increase performance. Note that, in Lindquist's (1995) study, both voice and vote include an explanation as to why standards must be higher than what is attainable by the participants in certain conditions. Table 1 presents a summary of variables, operationalizations, and key findings used in seven of the nine experimental research studies that have extended Lindquist (1995).

Libby (1999) operationalized voice as the opportunity to vote for preferred standards. She also tested explanation (yes/no) as to why the initial standard had to stay at a high amount. Cleary, Libby was actually testing the interaction of vote

¹ For more details on the model, see Vroom and Yetton (1973). *Leadership and Decision-making*. Pittsburgh: University of Pittsburgh Press.

Table 1 Operationalization of independent variables and key findings of prior research

Independent variables		Operationalization	Dependent variable	Key findings
Name	Values			
Lindquist (1995)	Voice	Vocalize feelings about standard setting process and standards received. In 2nd and 3rd periods also explanation for high standards and compromise in standard received	Budget Satisfaction Task Satisfaction Performance (no. of quality castles built)	Participants allowed only a voice in setting budgets experience greater budget and task satisfaction than those allowed no input, even with unattainable standards
	Vote	Opportunity to cast vote for a preferred standard with the knowledge it may or may not be accepted by management. In 2nd and 3rd periods also explanation for high standards and compromise in standard received		Vote is effective only when attainable budgets are received. When unattainable budgets are received, participants with a vote are less satisfied with the budget and the task. Participants form referent cognitions which lead to perceptions of pseudo-participation No effect for performance found
Libby (1999)	Voice	Opportunity to vote for preferred standard	Performance (no. of symbols correctly decoded)	Significant performance improvements with voice and explanation versus voice and no explanation
	Explanation	Explanation provided as to why standard must stay at original amount		However, really measuring impact of vote

Table 1 (continued)

Independent variables		Operationalization	Dependent variable	Key findings
Name	Values			
Chow et al. (2001)	Participation Treatment National Culture	Consultative: Opportunity to express an opinion on the standard setting process, the task and the preferred standard (i.e., combined voice and vote). Also giving the same explanation and compromise offered in Lindquist (1995) Autocratic: No chance for input of any kind; simply assigned performance standard each production period	Budget Satisfaction Task Satisfaction Adherence to Standard	US participants have significantly less satisfaction than Chinese with high-stretch (unattainable) standards Chinese participants adhere more strongly to imposed high-stretch targets
Libby (2003)	Type of Incentive Contact Contracting Process	Voice is operationalized through a case in which participants are told that their supervisor is sincerely interested in their opinions among other positive things about communication truth inducing / slack inducing fair / unfair	Budgetary Slack	Participants exposed to a fair contracting process (voice) produce less slack than participants exposed to an unfair contracting process with slack inducing incentive

Table 1 (continued)

Independent variables		Operationalization	Dependent variable	Key findings
Name	Values			
Bryne and Damon (2008)	Voice	Opportunity to communicate the preferred budget (i.e., vote)	Performance	Explanation group has higher performance than non-explanation group
	Explanation	Budget is or isn't greater than preference		Explanation type (different explanation each period) has a significant influence on performance
	Explanation type	Same or different explanation each period		Voice is measuring vote
	Trial number	Four periods		
Nahartyo (2013)	Type of procedural justice: Control Procedural Justice (CPJ)	CPJ is made up of 1. Voice as giving participants an opportunity to express their thoughts about initial budget 2. Vote as giving a chance to vote for final budget choice	Procedural Justice Judgments (PJJ) Budget Commitment	CPJ and GPJ both have independent positive impacts on PJJ There is also a significant interaction with the combination of CPJ and GPJ having the greater PJJ over all other combinations
	Type of procedural justice: Group Procedural Justice (GPJ)	GPJ contains telling positive things about the process in two aspects: 1. Integrity and honor are important to the supervisor 2. Budget setting process is neutral and workers have high status in the budgeting process and overall study		With stretch budgets, providing CPJ and GPJ results in the lowest Budget Commitment (frustration effect). Pseudo-participation is what they term using CPJ and GPJ with stretch targets

Table 1 (continued)

	Independent variables		Operationalization	Dependent variable	Key findings
	Name	Values			
Gomez-Ruiz and Rodriguez-Rivero (2018)	Consultative Participation	real high real low pseudo	Real High: Vote for preferred standard that is accepted Real Low: Vote for preferred standard but standard given is 20% higher Pseudo: Vote for preferred standard but given standard is much higher	Performance Autonomous Motivation Controlled Motivation	Autonomous motivation is higher when employees are in a real vs. pseudo consultative process Consultative participation exerts significant direct effect on autonomous motivation
	Perception of influence (control variable)		Scale		The positive effect of consultative participation on performance is mediated by autonomous motivation

and explanation, not voice. Measuring voice requires that participants vocalize their responses and not just write down their thoughts and feelings (Thibault and Walker 1975). So, in this case, Libby's (1999) operationalization of voice is actually Lindquist's (1995) representation of vote.

Chow et al. (2001) measured voice in the exact same manner as Lindquist (1995) but administered it in conjunction with vote to form consultative participation. They focused on national culture and found U.S. participants to have significantly less satisfaction than Chinese participants with high-stretch (unattainable) standards. Libby (2003) operationalized voice through a case where participants were told that their supervisor was sincerely interested in their opinions. She found exposure to voice in this fashion led to participants producing less slack in an experimental procedure. Byrne and Damon (2008) operationalized voice in the exact same fashion as Libby (1999) and thus also actually measured vote. Finally, Nahartyo (2013) operationalized control procedural justice as giving participants an opportunity to express their thoughts about an initial budget (voice). They additionally included voting for a preferred standard in this construct, much like Chow et al. (2001) did. He found controlled procedural justice to significantly boost perceptions of procedural justice.

As mentioned above, Libby (1999), in trying to measure voice, actually measured vote but did find significant performance improvements for participants allowed a voice (vote) and an explanation. Recall that Byrne and Damon (2008) also represented vote as voice and found no effect for voice. It seems the countervailing, pseudo-participation effect of vote with an unfavorable outcome (high standards) negated the positive impact of what they termed voice. Lastly, Gomez-Ruiz and Rodriguez-Rivero (2018) constructed consultative participation as involving either real high (vote for preferred standard and receive it), real low (vote for preferred standard and receive standard 20% higher), and pseudo-participation (vote for preferred standard and receive standard which is "much higher"). They found autonomous motivation to be higher when employees participate in a real versus a pseudo-consultative process.

The finding that vote for an attainable standard and subsequent receipt of an unattainable one results in perceptions of the vote representing pseudo-participation is corroborated by Francis-Gladney et al. (2008). They found, using a decoding experimental task, that, if participants receive a favorable (attainable) standard after voting for their preference, they believe budget participation was real. If, however, after a vote for a preferred (attainable) standard, participants receive an unattainable budget, pseudo-participation is blamed.

As discussed above, Lindquist did not find any performance effects when participants have a vote, that is, decision control, or both a voice and a vote, which represents the highest level of process control. Libby (1999), however, did find a positive performance effect for participants who received voice (vote) and an explanation for high standards versus participants with no voice (vote) or explanation. Chow et al. (2001) found Chinese participants adhered more strongly to stretch targets than did their U.S. counterparts. Byrne and Damon (2008) found higher performance for the group receiving an explanation for high performance standards. They additionally found offering a different explanation for high standards in each production round led to significantly higher performance than offering the same one

each time. While not directly measuring voice or vote, Francis-Gladney et al. (2008) found that, when participants received an unfavorable (unattainable) performance standard, they formed strong pseudo-participation perceptions. Also, Kelly et al. (2015) did not find any general impact of a form of procedural justice known as ex-post goal adjustments on performance. In their study, the adjustments are defined as an ex-post decrease in target budget, considering negative uncontrollable factors that arise during production and that enable a worker a better chance to meet goal and make bonus. Ex-post goal adjustments are an effective form of procedural justice when budgets received are moderate, but there is no impact on performance when standards are difficult.

The inconsistent findings from the studies above reveal that the relationship between procedural justice and satisfaction as well as performance in participative budgeting requires more reflection and clarification. In our study, we offer two main explanations for the contradicting results. First, as discussed above, we reason that conflicting findings stem largely from variations in research design. More precisely, the independent variable—procedural justice—has been operationalized and measured in different ways by Lindquist (1995) and subsequent researchers. Second, we propose that a manual-based task, as was used in Lindquist's original study and the follow-up study by Chow et al. (2001) that Lindquist co-authored, lead to different results than a mental-based decoding task used in all other follow-up experimental studies. We elaborate on this below.

2.3 Nature of task in experimental studies on procedural justice in participative budgeting

Environmental variables play a role in experimental research to elicit participants' responses in a pre-determined way (Birnberg and Nath 1968). Over 50 years ago, Birnberg and Nath (1968) laid the groundwork for experimental research in management accounting by recognizing the dichotomy between manual (physical) tasks versus those involving mental effort. They note that, in an iconic simulation like a laboratory experiment, the experimental task reduces a complex real-world manufacturing setting to a simpler substitute. They further indicate this substitute must be a representative task that participants react to in the same manner as manufacturing workers would to their tasks. Thus, the replication of a production line should employ a task that induces fatigue and boredom (Shin and Grant 2019). Boredom is defined as an unpleasant emotional state characterized by disinterest and difficulty with concentration (Fisher 1993; Loukidou et al. 2009).

Birnberg and Nath (1968) propose that the use of a mental task in a laboratory experiment can introduce an unintentionally high level of participant interest. When a task is intrinsically motivating, effort can be its own reward (Amabile 1993; Keller and Bless 2008). Chow's (1983) symbol decoding job is one such task. Since most experimental participants in the above-referenced research are students of accounting and other business studies, their perceptions of satisfaction and performance measures may be driven by their interest in the task at hand. They may develop an

intrinsic desire to improve upon their last performance, separate from the monetary payoff at the conclusion of the experiment (Shin and Grant 2019).

In all experimental studies extending Lindquist (1995), tasks are repeated a certain number of production periods. In Lindquist's (1995) study and the study by Chow et al. (2001), participants repeatedly build toy castles from the same number and kind of Loc Blocs© following an exact template. In the mental decoding tasks, participants are steadily confronted with a new set of symbols each production run and must translate these symbols into letters or words. Charness et al. (2018) find repetitions in constant manual tasks lead to fatigue and boredom of participants over time. While manual tasks are sometimes initially fun, they gradually become less motivating over time, leading to possible satisfaction and performance declines with repetitions. Thus Birnberg and Nath (1968) recommend manual tasks for research in management accounting, because these tasks better represent real-world situations, where the intrinsic interest of the job is low.

Two further task characteristics that are important within experimental accounting research are difficulty and the participant's familiarity with the task (Birnberg and Nath 1968). An ever-changing mental decoding task is likely more difficult than repetitive castle building, and more difficult tasks imply higher cost of effort (Charness et al. 2018). Further, most participants have presumably first-hand experience with Loc Blocs© but are less familiar with decoding tasks. When the level of difficulty or participant's familiarity with the task differ between tasks, results can be biased (Birnberg and Nath 1968). For example, in the relatively easy and well-known castle task, every participant should be able to build some castles with Loc Blocs©. Consequently, the level of performance should be quite similar across participants. In the more difficult and less familiar decoding task, some participants could excel while others fail. Given this, the level of performance might vary across participants.

The perception of task difficulty and the effort participants make to solve the respective task certainly depend also on participants' individual characteristics. In general, we may assume that solving mental tasks, similar to IQ tasks, depends mainly on participants' ability to make sense of complex facts, that is, the educative ability (Raven et al. 1998), and to store and reproduce information, that is, the reproductive ability (Raven et al. 1998). In decoding tasks, participants low in or without these abilities will likely be unable to compensate through greater effort. For example, participants who are uncomfortable with numbers or bad at memorizing, analyzing, or thinking creatively may have difficulty translating as many symbols into letters as more talented participants, no matter how hard they work. This is supported by Eckartz et al. (2012), who found that, in such cases, incentives seem to have very small effects on performance and differences in performance predominantly relate to individual skills rather than effort. Tasks like adding numbers or repetitive calculations depend clearly on both ability and effort (Eckartz et al. 2012). This might be the reason for variations in performance effects with manual versus mental experimental tasks. Clearly further research is needed to help clarify inconsistencies among performance effects in extensions of Lindquist (1995). We believe comparable results can be achieved by only replicating Lindquist's research design as exactly as possible, including using a manual task, as he did.

2.4 Hypotheses

If our assumptions are correct and inconsistent findings in follow-up studies stem from variations in research design, a replication study that retains Lindquist's research design and employs a manual-based experimental task should produce the same results. We embrace calls to replicate key studies before branching into new extensions of the work (Otley and Pollanen 2000). These calls also suggest slight variations in measures or empirical methods might add interest to the replication. Thus we hold constant Lindquist's (1995) manipulations of procedural justice (i.e., voice and vote) and focus only on situations of distributive injustice or receipt of unattainable standards. We also introduce a new manual-based task (i.e., paper quilt making) to the literature to illustrate that subsequent findings can be robust when retested with consistent experimental metrics. We therefor offer three hypotheses regarding the impact of procedural justice on satisfaction and performance, as did Lindquist (1995).

Hypothesis 1 *Individuals allowed a voice only will be more satisfied with their budgets than individuals with similar ability offered no input or a vote only when unattainable budgets are received.*

Hypothesis 2 *Individuals allowed a voice only will be more satisfied with the experimental task than individuals with similar ability offered no input or a vote only when unattainable budgets are received.*

Hypothesis 3 *Individuals allowed a voice only will perform better than individuals with similar ability offered no input or a vote only when unattainable budgets are received.*

3 Research method

3.1 Participants

Our population sample consists of 76 undergraduates from a mid-sized state university in Southern Austria.² To recruit participants, an email was sent to all enrolled students at the university. Students were invited to register anonymously for time slots provided in Doodle and arrived at rooms prepared for the experimental study at their scheduled time. The sample was first reduced to 69 participants, because the time limit of one hour was exceeded in the first experimental run with six parallel sessions and because one participant in another session did

² The experimental materials of Lindquist's (1995) original study, which was conducted with an English-speaking population, were translated into German by one of the researchers fluent in German and English and confirmed by an unrelated third-party in Europe, also fluent in both languages. Copies of full experimental materials are available in the appendices.

not understand the incentive scheme properly. The sample was again reduced to $n=66$ after a proper yoking of participants occurred to balance conditions on ability.³ The mean age of participants in our experiment is 25.2 years, and the participants include 22 men and 47 women. Their full-time work experience is 7.4 years and part-time 2.7 years on average. Thirty-six percent of participants come from business and management, 28% from cultural studies (e.g., language, teaching and education), 26% from psychology, and 10% from technical studies. We tested for confounding effects due to age, work experience, and field of studies. No effects were found.

3.2 Experimental task

Recall Lindquist (1995) had participants build toy castles from Loc Blocs®. Castles were constructed of 26 plastic toy pieces of various sizes and shapes. A quality castle needed to match a model in size and number of pieces. Additionally, four key pieces needed to match in size and exact color of pieces. Our task is designed to emulate the manual aspects of the castle building. Participants are charged with constructing paper quilts out of pre-cut scraps by gluing the paper scraps down on a sheet of paper, following a model of a quality quilt given to them. Figure 1 presents the model of a quality quilt.

Participants are provided with a box filled with pre-cut pieces of paper in assorted colors, shapes, and sizes. They are also provided a glue stick and a stack of white paper on which is printed an outline grid (i.e., an uncolored, blank rectangle), showing where to place pieces. The outline grid measures 11 cm × 10 cm. As the model of a quality quilt in Fig. 1 shows, the pieces no. 3 and no. 6 are of the same size and shape. All other pieces have different sizes and shapes. Only six paper pieces are needed to complete one quality quilt, but they must all fit the grid perfectly in size and shape. Also, two of the pieces (i.e., no. 3 and no. 5) need to match not only in size and shape but color as well (i.e., yellow and blue, respectively). Furthermore, one piece (i.e., no. 6) must not be yellow. We deliberately use light pastel paper of six different colors so that colors were not readily identifiable. We include a question in our final survey to ensure none of our participants were colorblind.

³ It was further necessary to match participants on ex-post preferability of budgets (i.e., matched subjects would both have to consider standards as either attainable (fair) or unattainable (unfair)). This was accomplished by matching on at least one number of each participant's attainable range (e.g., if a vote subject's range was 8, 9, 10 it was considered a match to a no vote subject with a range of (7, 8, 9), (8, 9, 10) or (10, 11, 12)). In 48 percent of cases, these numbers matched perfectly in all three periods. Eighty-eight percent matched perfectly in at least two periods, and 97 percent matched perfectly in at least one period.

Fig. 1 Model of the paper quilt

3.3 Independent variables

We use a 2×2 full-factorial design with two independent, between participants variables to operationalize procedural justice: (1) voice and (2) vote.⁴ Unlike Lindquist (1995), we do not manipulate standard attainability (fairness), because we are only interested in the outcomes related to unattainable (unfair) standards. When standards set are more difficult to achieve than the ones attainable by the subordinates (i.e., the participants), the budget allocation is perceived to be unfair, and this helps to ensure that subordinates concentrate on the fairness of the process in making their overall fairness judgments. (See also Byrne and Damon 2008; Libby 1999 based on the two-component model of justice by Cropanzano and Folger 1991.)

Our manipulations of voice and vote exactly follow Lindquist (1995). Voice is manipulated as either giving participants the opportunity to express their opinions, thoughts, feelings, etc., (voice) or not (no voice) before receiving their standards. Following Lindquist (1995), voice also includes an explanation for receiving unattainable standards as well as a compromise in the amount of product that must be built in the experiment in periods two and three (which are the

⁴ We recognize that, since our hypotheses do not call for analyses of main effects or interactions, we could view our analysis as a one-way ANOVA (1×4 design) with one independent variable measuring four levels of process fairness. We choose to analyze our study as a 2×2 , since it is a replication of Lindquist (1995) and he designed his original study measuring voice and vote as separate independent variables. Supplementary statistical analyses of our results as one-way ANOVAs with planned comparisons, however, do support our findings.

VOICE		NO VOICE	
VOTE 1	NO VOTE 2	VOTE 3	NO VOTE 4

Fig. 2 Experimental conditions

periods where standards are ratcheted up 30% and 70% higher than what is attainable). Details of the compromise are discussed below.

When participants are given a voice, the experimenters, that is, both the foreman and manager, encourage the participants to comment on the standard setting and on the level of standards set in previous periods. They also ask for the range of standards that the participants think is attainable. By contrast, when participants do not have a voice, they are not granted any of these possibilities and communication between participants and experimenters is minimized.

Participants in the vote condition alone are never asked what range of standards they feel is attainable. They write that information down for themselves, but the information is kept private. In the vote manipulation, participants are either given the right to communicate their preferred standard (vote) or they are not asked for their preference (no vote). When participants are given a vote, they are told that a person's personal choice influences the management's decisions, even if management has the final say. Participants with a vote are reminded to make sure to choose standards that maximize their compensation within their ability. Participants receiving a vote are also given an explanation and compromise as discussed above. By contrast, when participants do not have a vote, they receive standards set solely by management, regardless of what they think is attainable (unless they have a voice). Consequently, there are four possible conditions. Figure 2 illustrates them.

Irrespective of vote, individuals given a voice have low process control (VOICE in Fig. 2), while individuals without a voice have less process control (NO VOICE in Fig. 2). Similarly, irrespective of voice, individuals who are offered a vote have high process control (VOTE in Fig. 2), while individuals without a vote have low process control (NO VOTE in Fig. 2). Consequently, individuals who are given both voice and vote have the highest level of process control, while individuals with neither a voice nor a vote have no input at all. Individuals in the no-input condition rely solely on the incentive contract to motivate their performance.

The independent variable voice is validated with an ANOVA measuring responses to two statements from the final questionnaire on a Five-point Likert scale, which are combined into one construct. A higher mean indicates a stronger perception of having a voice. The independent variable vote is again validated with an ANOVA measuring the response to one statement from the final questionnaire, which is also on a Five-point Likert scale with a higher mean indicating a stronger perception of having a vote. Further details on the measurement as well

as results are presented below in the section on the manipulation checks. Table 2 provides an overview of all variables, their definitions, values, and measurements.

3.4 Dependent variables

Three dependent variables are measured in our study: (1) satisfaction with unattainable budgets received, (2) satisfaction with the experimental task, and (3) performance. Each of the two dependent variables on satisfaction is measured with responses to four statements from the final questionnaire on a Five-point Likert scale, which are combined into one construct. A higher mean indicates greater satisfaction. Performance is measured as the summation of actual production of quality paper quilts in production rounds 2 and 3, because these are the two periods where standards are ratcheted up 30% and 70% respectively, either from participants' preferred standards when they are given a vote (cond. 1 and 3) or from participants' vocalized attainable standards when they are given a voice but no vote (cond. 2) or from participants' supposedly only privately noted attainable standards when they cannot give any input (cond. 4). Details regarding standard attainability are presented in Sect. 3.5, and the experimental pacing is described in detail in Sect. 3.7.

3.5 Measurement of standard attainability across conditions

The written range of attainable standards requested at the beginning of each production period for all participants is private information. Use of this data is contingent upon condition. For condition 1 (voice and vote), participants first orally discuss the attainable range they wrote down and express feelings about it (voice). Then they are asked to vote for a preferred standard by which they will work under the truth-inducing incentive scheme. A performance standard is determined to be the chosen standard in production period 1, and then one set 30% higher than chosen in period 2 (compromising from management's initial desire to increase the chosen standard by 45%), and finally 70% higher than chosen in period 3 (compromising from management's initial desire to increase the chosen standard by 85%). Thus the written attainable range is not used. In condition 2 (voice/no vote), participants speak about their attainable range but are not given a chance to vote for a preference. Thus their standard is determined as the midpoint in period 1, 30% above the midpoint in period 2 (after the same compromise from 45% discussed above), and 70% above the midpoint in period 3 (after the same compromise from 85% also discussed above). Again the written attainable range isn't used.

In condition 3 (no voice/vote), standard setting follows the pattern of basing the budget off the chosen standard as in condition 1, with the compromises in periods 2 and 3, again ignoring the written attainable range. Finally in condition 4 (no voice/no vote), the written attainable standard range is used to set standards. Even though the information was deemed to be private, experimenters in actuality could glance at the written attainable range, before participants set it aside, and establish a midpoint, which they expressed to management. Management then sets the standard at

Table 2 Definition of variables in tables

Variable	Description of variable	Value range	Statements in the questionnaire	Scale	Answer options
VOICE	Procedural justice by having a voice as a construct of 2 statements	$2 \leq \text{VOICE} \leq 10$ with a higher value indicating higher procedural justice	"I could vocalize my feelings."	5-Point Likert	(1) strongly disagree (5) strongly agree
			"I had no chance to tell the foreman how I felt." (<i>reverse-scored</i>)	5-Point Likert	(5) strongly disagree (1) strongly agree
VOTE	Procedural justice by having a vote	$1 \leq \text{VOTE} \leq 5$ with a higher value indicating higher procedural justice	"My decision as to the level of my standard had a lot to do with the standard that had been set for me."	5-Point Likert	(1) strongly disagree (5) strongly agree
			"On the whole, I was satisfied with the standards under which I worked."	5-Point Likert	(1) strongly disagree (5) strongly agree
BUDSAT	Satisfaction with unattainable budgets received as a construct of 4 statements	$4 \leq \text{BUDSAT} \leq 20$ with a higher value indicating higher satisfaction	"The standards I worked under were convenient for me."	5-Point Likert	(1) strongly disagree (5) strongly agree
			"I would have preferred standards different from those I received." (<i>reverse-scored</i>)	5-Point Likert	(5) strongly disagree (1) strongly agree
			"I liked the standards which I received."	5-Point Likert	(1) strongly disagree (5) strongly agree

Table 2 (continued)

Variable	Description of variable	Value range	Statements in the questionnaire	Scale	Answer options
TASKSAT	Satisfaction with the experiential task as a construct of 4 statements	$4 \leq \text{TASKSAT} \leq 20$ with a higher value indicating higher satisfaction	<p>"On the whole, I was satisfied with this task."</p> <p>"I would recommend this task to someone else as one that is satisfying."</p> <p>"I found this task to be unpleasant." (<i>reverse-scored</i>)</p> <p>"I feel fortunate to have been involved in this task."</p>	<p>5-Point Likert</p> <p>5-Point Likert</p> <p>5-Point Likert</p> <p>5-Point Likert</p>	<p>(1) strongly disagree (5) strongly agree</p> <p>(1) strongly disagree (5) strongly agree</p> <p>(5) strongly disagree (1) strongly agree</p> <p>(1) strongly disagree (5) strongly agree</p>
Performance	Actual production of	$0 \leq \text{Performance} \leq 14$	Summation of quality paper quilts produced in production rounds 2 and 3	Metric	
TOTAL-FAIR	Perception of fairness across production rounds 1–3	$3 \leq \text{Performance} \leq 15$ with a higher value indicating a higher perception of fairness	<p>In every production round:</p> <p>"The way this company goes about setting my standard is"</p>	5-Point Likert	<p>(1) extremely unfair (5) extremely fair</p>

the midpoint for period 1, 30% above the midpoint for period 2, and 70% above the midpoint for period 3 with no explanation.

3.6 Compensation

Participants are paid a piece-rate salary based on quality units produced. Pre-testing determined that € 0.40 was appropriate for our study. Payment is calculated according to the scheme in the following equation.

$$Y = € 0.40(A) - € 0.40|A - S|,$$

where Y is the individual pay per production run, S is the performance budget, and A is the actual performance of the individual. This truth-inducing incentive scheme motivates individuals to attain budgets or face financial penalty. It also motivates them to express as high a budget as they feel is attainable to maximize their financial rewards, as there is no financial reward for producing over standard. This scheme is effectively employed in previous accounting research (Chow et al. 2001; Lindquist 1995; Young 1985; Young et al. 1993).

Experimental participants are walked through the following examples until they confirm they understand the scheme. Imagine a standard is 6, and actual production is 4. Financial compensation is then € 0.40 (4)–€ 0.40 |4–6|=€ 0.80. That means compensation is reduced by € 0.80 penalty for not obtaining the standard. If, however, a standard of 6 matches actual production of 6, then financial compensation is € 0.40 (6)–€ 0.40 |6–6|=€ 2.40. Finally, if standard is 6 and actual production is 8, participants do not receive any additional compensation but only € 0.40 (8)–€ 0.40 |8–6|=€ 2.40. Thus there is no financial reward for producing beyond standard. Overall, participants receive higher compensation when they make more quality quilts but only if they do not deviate from the standard, that is, the performance budget.

Average pay to participants is € 10 for one hour of their time. Our experiment ran half the time of Lindquist's (1995), which was necessary due to time constraints. We shortened the production runs from 10 min to eight and generally kept a brisk pace to make up time. Also, the paper quilt task, with only six pieces of assemblage, did not take as long to complete as the toy castle task, with 26 pieces.

3.7 Experimental pacing

Figure 3 presents detailed, side-by-side steps of our experiment, as compared to Lindquist (1995). In every phase of the process, except for administering the paper quilt task, instead of the castle building task, we replicate Lindquist's work as closely as possible. Participants signed up for a one-hour block of time and, upon arrival, were greeted by the head experimenter, playing the role of the manager, and were escorted to a seating area to await production room assignments. Production rooms were sealed off from one another regarding sight and sound. Participants were unaware of what their "co-workers" were experiencing. In all conditions, experimenters followed a written script (see Appendices B and C).

Experiment Steps in Lindquist (1995) versus Our Study (2020)

PARTICIPANTS	Lindquist (1995)	Our Study (2020)
Number of full-time undergraduates	111	76
Reduced number after yoking on ability	86	66
Mean age	21.6	25.2
Gender	43 men, 43 women	22 men, 47 women
Mean part-time work experience	1.84 years	7.4 years
Mean full-time work experience	4.26 years	2.70 years
Average pay:	\$ 10.00 for 2 hours	€ 10.00 for one hour
EXPERIMENTAL TASK	Lindquist (1995)	Our Study (2020)
Kind of manual task	Building toy castles from Loc Blocs®	Making paper quilts from paper scraps
Task requirements	<ul style="list-style-type: none"> • 26 blocks or 6 types of paper scraps, respectively, of varied sizes, shapes, and colors • Quality castles or quilts, respectively, needed to match the model in size, color, and number of pieces • 4 or 2 key pieces, respectively, needed to match in size and color of pieces 	
INDEPENDENT VARIABLES	Lindquist (1995)	Our Study (2020)
Full-factorial design	2x2x2	2x2
Number of independent variables	Three	Two
Independent variables	<ol style="list-style-type: none"> 1. Procedural justice – Voice with voice as low process control versus no voice as no process control 2. Procedural justice - Vote with vote as high process control versus no vote as no process control 3. Distributive justice with attainable (fair) versus unattainable (unfair) budgets 	
		All participants receive unattainable (unfair) budgets.
DEPENDENT VARIABLES	Lindquist (1995)	Our Study (2020)
Number of dependent variables	Three	
Dependent variables	<ol style="list-style-type: none"> 1. Budget satisfaction - measured on Five-point Likert 2. Task satisfaction - measured on Five-point Likert 3. Performance - measured as the actual number of quality units produced across all three production periods or production periods two and three, respectively. 	
EXPERIMENTAL PROCEDURES	Lindquist (1995)	Our Study (2020)
Duration of one session	Participants sign up for one two-hour or one-hour block of time, including 20 or 16 minutes of training, respectively, and three production periods.	
Duration of the experimental study	One week to complete running from 8:00 am to 7:30 pm (Monday – Friday)	Two days to complete running from 8:00 am to 5:00 pm (Tuesday and Wednesday)
Yoking process	Done to balance ability across conditions	
Truth-inducing incentive scheme	Used to reduce the propensity to create slack	
Assignment to conditions	Participants are randomly assigned to the conditions.	

Fig. 3 Experiment steps in Lindquist (1995) versus our study (2020)

We used six experimenters who were familiar with all four experimental conditions, but each experimenter was always assigned the same two conditions of the four. The manager was played by the same person in all sessions. Experimenters underwent approximately one hour and a half of training before the experiment began. The manager led each participant to one of the production rooms and introduced him or her to the foreman (i.e., the experimenter). Next the foremen read instructions and the case scenario to the participants and had them complete an unpaid practice period of eight minutes followed by an additional eight-minute production period with a set payment of € 0.40 per paper quilt. The objective of the unpaid practice round was for participants to familiarize themselves with the task,

EXPERIMENTAL PROCEDURES (continued)	Lindquist (1995)	Our Study (2020)
Number of conditions	Eight	Four
Conditions		
(1) Voice/Vote/Attainable Budget	yes	n/a
(2) Voice/No Vote/Attainable Budget	yes	n/a
(3) No Voice/Vote/Attainable Budget	yes	n/a
(4) No Voice/No Vote/Attainable Budget	yes	n/a
(5) Voice/Vote/Unattainable Budget	yes	yes
(6) Voice/No Vote/Unattainable Budget	yes	yes
(7) No Voice/Vote/Unattainable Budget	yes	yes
(8) No Voice/No Vote/Unattainable Budget	yes	yes
Reception of participants	Participants, approximately five at a time, were greeted by the head experimenter playing the role of the manager and escorted to a seating area to await production room assignments.	
Preparation of production rooms	Production rooms were completely sealed off from one another regarding both sight and sound. Thus, participants were unaware of what their co-workers were experiencing. There was one experimenter per room playing the role of the foreman. Each production room contained two chairs and a table on which sat a box filled with Loc Blocs © pieces or paper quilt pieces (plus glue stick and white paper) , respectively, of assorted colors and sizes. In all cases, experimenters followed a written script.	
Experimenter training before experiment	two hours	one and a half hours
Start of Experiment	Participants received the scenario that described the task at hand.	
Practice periods	<ul style="list-style-type: none"> • 10/8 minutes unpaid • 10/8 minutes paid with \$/€ 0.40 per unit 	
Production period ONE for Condition 1 as example	<ul style="list-style-type: none"> • Range attainment setting 1 • Voice 1 • Vote 1 • Budget Setting 1 • 1st official production run (10 or 8 minutes, respectively) • Count quality toy castles or quality paper quilts, respectively, and ready for the next run. 	
Production period TWO for Condition 1 as example	<ul style="list-style-type: none"> • Range attainment setting 2 • Voice 2 • Vote 2 • Budget Setting 2 • 2nd official production run (10 or 8 minutes, respectively) • Count quality toy castles or quality paper quilts, respectively, and ready for the next run. 	
Production period THREE for Condition 1 as example	<ul style="list-style-type: none"> • Range attainment setting 3 • Voice 3 • Vote 3 • Budget Setting 3 • 3rd official production run (10 or 8 minutes, respectively) • Count quality toy castles or quality paper quilts, respectively. 	
End of the experimental session for all conditions	<ul style="list-style-type: none"> • Final Questionnaire • Payment and Debriefing • Exit 	

Fig. 3 (continued)

ask for assistance if necessary, and get detailed feedback on the quality of their quilts. The objective of the paid practice round was to familiarize participants with a piece-rate incentive of € 0.40 per paper quilt. Next participants completed three actual production runs of eight minutes each using the truth-inducing incentive scheme. To avoid potential end-of-game effects, participants were not told the total number of production periods. Participants were placed randomly into one of four

conditions; (1) voice and vote, (2) voice only, (3) vote only, (4) no input. The experiment took two days to run from 8:00 am to 5:00 pm both days.

In all conditions, participants were asked to complete a series of documents. First, there was a questionnaire on demographics at the beginning of the experiment. Second, at the beginning of each of the three production rounds, participants were asked to state their attainable range of production of quality paper quilts in one production period of eight minutes, to obtain their individual ability measures and to have a baseline for the increase in standards in the no-input condition (no voice and no vote). This questionnaire was put aside by the foreman so that participants perceived the attainable range as private information at this point. However, in all cases, the foreman could glance at this attainable range and relay that information to the manager. The third questionnaire was administered immediately following the setting of performance standards in every production period and asked participants to respond to the following statement: "The way performance standards are set around here is fair." Its purpose was to measure participants' perceptions of fairness in the sense of procedural justice as they moved through the experiment. This document was also placed away from the foreman and represented private information to the participant. Finally, after the third and last production period, all participants filled out a final survey questionnaire, which provided data for dependent measures related to satisfaction as well as other information (e.g., for manipulation checks).

4 Results

4.1 Demographics

Even though participants were randomly assigned to experimental conditions, an ANOVA of demographic data is conducted to ensure they were evenly distributed. Data includes age, gender, and full-time or part-time work experience. There are no significant main effects or two-way interactions. Chi square tests are conducted to ensure the even distribution. There are no significant differences between conditions regarding age, gender, and work experience.

4.2 Manipulation checks

Manipulation checks involve checks on the independent variables of voice and vote (procedural justice). ANOVAs are conducted to test the manipulations. We conduct attention manipulation checks to measure whether voice and vote were perceived as existing. The first manipulation check for voice aims to determine whether voice is viewed as participation (procedural justice). Responses to two different statements in the final questionnaire are summed to measure differences in opinion regarding the amount of voice allowed participants. The first statement reads: "I could vocalize my feelings." The second states: "I had no chance to tell the foreman how I felt" (reverse-scored). Scale reliability for this two-item scale is very satisfactory, with a Spearman-Brown coefficient of 0.81. An ANOVA of the full model, with

Table 3 Means for manipulation checks and hypotheses

	Voice and vote	Voice only	Vote only	No input	Voice	No voice	Vote	No vote
<i>Manipulation check 1</i>								
Voice	8.00	8.28	4.24	6.28	8.14	5.25	6.12	7.27
<i>Manipulation check 2</i>								
Vote	2.81	3.56	3.12	2.87	3.19	3.00	2.97	3.21
<i>Support for pseudo-participation from vote</i>								
Totalfair	9.19	10.33	8.35	7.67	9.76	8.01	8.76	9.12
<i>Hypothesis 1</i>								
Satisfaction with budget	11.19	12.06	9.25	10.87	11.62	10.06	10.22	11.46
<i>Hypothesis 2</i>								
Satisfaction with task	13.00	13.89	11.65	12.40	13.44	12.06	12.32	13.18
<i>Hypothesis 3</i>								
Performance	5.44	5.72	3.65	4.20	5.58	3.92	4.52	5.03

the composite dependent variables is run. A strong main effect is found for voice ($F_{1,65}=27.08$, $p<0.00$; voice 8.14 versus no voice 5.25). Investigation of two-way t-test contrasts indicates the manipulation of voice is robust across vote conditions. Participants with a voice only (8.28) significantly believe they have more voice than participants allowed a vote only (4.24); $t_{33}=5.14$, $p<0.00$. These findings lend strong support that the voice manipulation worked as intended. Thus voice is viewed as a means of participation. Tables 3, 4, 5 provide means, ANOVA results, and t-test contrasts for the manipulation checks and tests of hypotheses.

Note that a main effect for vote is also significant in the two-way ANOVA ($F_{1,65}=4.33$, $p<0.042$; vote 6.12 versus no vote 7.27). One might expect this to happen, as participants with a vote are asked for their preference of standard. However, this significance indicates participants with no input (no voice and no vote) perceive they have significantly greater voice than those with a vote only (6.28 versus 4.24). This phenomenon will be discussed in upcoming sections.

The second manipulation check for vote determines whether vote is viewed as participation (procedural justice). Responses to the statement, "My decision as to the level of my standard had a lot to do with the standard that had been set for me," are used as a dependent variable in an ANOVA of the full model. A significant main effect is not found for vote ($F_{1,65}=0.65$, $p<0.42$; vote 2.97 versus no vote 3.21). This is not too surprising, since our hypotheses claim that voice will have the strongest positive impact on the dependent measures, as compared to vote or a combination of voice and vote. We do find,⁵ however, that the two-way interaction is almost significant at $F_{1,65}=2.67$, $p<0.11$, which calls for t-test contrasts.

⁵ All t-test contrasts are presented as one-tailed tests with the exception of manipulation checks which are two-tailed.

Table 4 ANOVA results for effects of voice and vote on fairness perception, budget satisfaction, task satisfaction and performance

	Voice	Vote	Voice × Vote
<i>Manipulation check 1: voice</i>			
<i>F</i>	27.08	4.33	2.50
Probability	0.00**	0.04**	0.12
<i>Manipulation check 2: vote</i>			
<i>F</i>	0.40	0.65	2.67
Probability	0.53	0.42	0.11
<i>Support for pseudo-participation from vote: totalfair</i>			
<i>F</i>	5.56	0.10	1.52
Probability	0.02**	0.76	0.11
<i>Hypothesis 1: satisfaction with budget</i>			
<i>F</i>	2.14	1.35	0.12
Probability	0.15	0.25	0.72
<i>Hypothesis 2: satisfaction with task</i>			
<i>F</i>	2.21	0.74	0.01
Probability	0.14	0.39	0.94
<i>Hypothesis 3: performance</i>			
<i>F</i>	4.97	0.32	0.32
Probability	0.03**	0.58	0.86

The vote manipulation should not result in a statistically significant difference, since vote is seen as a form of pseudo-participation. As we predict, voice is the stronger form of procedural justice in our study. This is confirmed by comparing participants with a voice and a vote (2.81) to those with a voice only (3.56); $t_{32} = -1.72$, $p < 0.10$. Individuals with a voice only perceive they have more voice than those with a voice and a vote. This supports vote being seen as a form of pseudo-participation. Note that these findings do not imply a vote is not a form of high process control, but rather, as in Lindquist (1995), it is just not seen as one when unattainable work standards are assigned.

4.3 Tests of hypotheses

To test all hypotheses, full-factorial models are analyzed separately with dependent variables of budget satisfaction, task satisfaction, and performance. Budget satisfaction reflects individuals' satisfaction with unattainable budgets (i.e., standards) received and is measured with responses to four statements from the final questionnaire. The four statements are the following. 1) On the whole, I was satisfied with the standards under which I worked. 2) The standards I worked under were convenient for me. 3) I would have preferred standards different from those I received

Table 5 T-test results for effects of voice and vote on fairness perception, budget satisfaction, task satisfaction and performance (one-tailed)

	Voice and vote Voice only	Voice and vote Vote only	Voice and vote No input	Voice only Vote only	Voice only No input	Vote only No input
<i>Manipulation check 1: voice (two-tailed)</i>						
<i>t</i>	−0.38	4.86	2.24	5.14	2.53	−2.42
Probability	0.70	0.00**	0.04**	0.00**	0.02**	0.02**
<i>Manipulation check 2: vote (two-tailed)</i>						
<i>t</i>	−1.72	−0.77	−0.12	1.05	1.49	0.59
Probability	0.10**	0.46	0.90	0.30	0.14	0.56
<i>Support for pseudo-participation from vote: totalfair (one-tailed)</i>						
<i>t</i>	−1.27	0.84	1.30	2.12	2.42	0.57
Probability	0.11	0.20	0.11	0.02**	0.01**	0.29
<i>Hypothesis 1: satisfaction with budget (one-tailed)</i>						
<i>t</i>	−0.64	1.24	0.19	2.10	0.81	−0.97
Probability	0.27	0.11	0.43	0.02**	0.21	0.17
<i>Hypothesis 2: satisfaction with task (one-tailed)</i>						
<i>t</i>	−0.71	1.00	0.41	1.78	1.11	−0.52
Probability	0.24	0.16	0.34	0.04**	0.14	0.31
<i>Hypothesis 3: performance (one-tailed)</i>						
<i>t</i>	−0.30	1.62	1.12	2.08	1.53	−0.50
Probability	0.39	0.06*	0.14	0.03**	0.07*	0.31

(reverse-scored). 4) I liked the standards which I received. Responses to these four statements are measured on a Five-point Likert scale and are combined into one construct named BUDSAT with a Cronbach's alpha of 0.91.⁶ The higher BUDSAT, the more satisfied individuals are with the budget, that is, the standards they receive.

Task satisfaction reflects individuals' satisfaction with the experimental task and is measured by the summation of responses to four statements from the final questionnaire. All four statements are measured on Five-point Likert scales, too. They are the following. (1) On the whole, I was satisfied with this task. (2) I would recommend this task to someone else as one that is satisfying. (3) I found this task to be unpleasant (reverse-scored). (4) I feel fortunate to have been involved in this task. Responses from these four statements are combined into one construct called TASKSAT with a Cronbach's alpha of 0.80. The higher TASKSAT, the more satisfied individuals are with their experimental task.

Performance is measured as the summation of actual production of quality paper quilts in production rounds two and three, where standards are ratcheted up 30% and 70% respectively from the midpoint of participants' preferred ranges of standards (cond. 1 and 3), vocalized attainable ranges (cond. 2), or attainable ranges discreetly communicated to the manager (cond. 4), depending on the condition. Performance and standards set in each period are reported in Table 6.

⁶ Separate factor analyses were run for each of the satisfaction constructs and determined that all statements in each construct loaded on one factor.

Table 6 Means for standards set and performance per period

	Voice and vote	Voice only	Vote only	No input
<i>Standards set per condition and period (mean)</i>				
In period 1	2.63	2.72	2.18	2.53
In period 2	3.44	3.50	2.94	3.20
In period 3	4.81	4.50	4.47	4.33
<i>Performance per condition and period (mean)</i>				
In period 1	1.87	1.56	1.12	1.40
In period 2	2.38	2.44	1.59	1.93
In period 3	3.06	3.28	2.06	2.27
In periods 2 and 3	5.44	5.72	3.65	4.42

4.3.1 Hypothesis 1

Hypothesis 1 predicts that individuals allowed a voice only will be more satisfied with standards received than individuals with similar ability offered no input or a vote only, when unattainable budgets are received. The main effect for voice is not significant; $F_{1,63} = 2.14$, $p < 0.15$. However, we believe this finding is driven by the negative influence of vote (i.e., half of the participants with a voice also had a vote). Lindquist (1995) established that having a vote, though it is a higher level of process control, results in satisfaction declines with the budget. Therefore the real test of this hypothesis lies in the contrast analysis of participants. Here we find participants with a voice only (12.06) are more satisfied with stretch targets than those with no input (10.87), but the difference isn't significant; $t_{31} = 0.81$, $p < 0.21$. More importantly, we also find participants with a voice only (12.06) are significantly more satisfied with stretch targets, as compared to those with a vote only (9.25); $t_{31} = 2.10$, $p < 0.02$. Overall, we find support for Lindquist's (1995) original findings regarding levels of procedural justice and budget satisfaction. Hypothesis 1 is partially supported.

Two questions then arise: why did the influence of vote drive the main effect for voice above to insignificance, and why does voice lead to significantly greater satisfaction with unattainable standards than vote? The reason lies in participants' perception that vote is an insincere form of procedural justice or a form of pseudo-participation. To address why this might be, we examine the individuals' perception of fairness through the entire experiment. To measure this effect, we create a new variable, TOTALFAIR, by summing the fairness perceptions in production rounds 1–3. After each round, participants are presented with this statement: "The way standards are set around here is fair." A Five-point Likert scale is used, with a higher level indicating higher perception of fairness. A full factorial ANOVA is then run with the independent variables of voice and vote. As expected, a main effect for voice is significant at $F_{1,65} = 5.56$, $p < 0.02$; voice (9.76) versus no voice (8.01). The main effect for vote, however, is not significant at $F_{1,65} = 0.10$, $p < 0.76$. Contrast analysis confirms this main effect for voice and finds participants with a voice only (10.33) have significantly higher perceptions of process fairness over those with no input (7.67) at $t_{31} = 2.42$, $p < 0.01$. Participants with a voice only (10.33) also perceive

significantly higher budget setting fairness than do those with a vote only (8.35) at $t_{33}=2.12$, $p<0.02$. A perception of pseudo-participation seems to arise when a participation process is deemed unfair.

4.3.2 Hypothesis 2

Hypothesis 2 predicts that individuals allowed a voice only are more satisfied with the experimental task than individuals with similar ability offered no input or a vote only, when unattainable budgets are received. An ANOVA of the full model finds both main effects for voice and for vote as well as their interaction to be insignificant. Contrast analyses further find participants with a voice only (13.89) to have greater task satisfaction than those with no input (12.40), but again the difference is not significant; $t_{33}=1.11$, $p<0.14$. Again we also find participants with a voice only (13.89) are more satisfied with the experimental task than those with only a vote (11.65) at $t_{33}=1.78$, $p<0.04$. It seems again that voice's insignificant main effect is being reduced by the half of that population that also had a vote. Perceptions of unfairness related to vote again result in a higher form of process control undercutting satisfaction with an experimental task. We find partial support for hypothesis 2.

4.3.3 Hypothesis 3

Hypothesis 3 predicts that individuals allowed a voice only will perform better than individuals with similar ability offered no input or a vote only when unattainable budgets are received. The ANOVA of the full factorial model shows a significant main effect for voice at $F_{1,65}=4.97$, $p<0.03$. Participants with a voice (5.58) have significantly greater performance, compared to those with no voice (3.92). Unsurprisingly, contrast analysis indicates that participants with a voice only (5.72) significantly outperform those with no input (4.20) at $t_{31}=1.53$, $p<0.07$. In line with other findings, participants with a voice only (5.72) also outperform those with only a vote (3.65) at $t_{33}=2.08$, $p<0.03$. Hypothesis 3 is fully supported.

5 Discussion and conclusions

We designed this study to reconsider some of the findings regarding the impact of procedural justice on satisfaction and performance that have arisen since publication of Lindquist's (1995) seminal study. Lindquist (1995) manipulated two forms (low/high) of procedural justice as voice and vote. Voice allowed participants to express their feelings regarding the setting of a piece-rate budget. Vote allowed them to vote for a preferred standard. Lindquist found that a lower form of process control in setting standards (voice) led to enhanced satisfaction over a higher form of process control (vote) when one would have expected the opposite. Subsequent partial replications and extensions, however, have found relationships among voice, vote, and varied dependent measures to be in partial agreement and

in opposition to Lindquist's findings. Some research even finds voice to be a form of pseudo-participation.

We contend there are two primary problems with the formation of experimental metrics in this prior work. First, subsequent research employs differing operationalizations of procedural justice, as compared to Lindquist. Additionally, the preponderance of research trying to extend Lindquist (1995) has employed a mental-based experimental task (symbol decoding), as contrasted with Lindquist's manual-based task (castle building), even though previous research suggests manual, repetitive tasks are necessary to properly represent a manufacturing environment. It is questionable whether mental-based tasks approximate manufacturing workers' reactions to measures of procedural justice.

Our study replicates Lindquist (1995) by employing exact operationalizations of his measures of voice and vote. We even obtained his scripts and experimental documentation for the abovementioned four conditions and translated them to German for the experiment to be run in Europe. We also purposely administered a different task—paper quilt making—to test whether that could elicit findings similar to those of Lindquist. We find, as did Lindquist, that participants allowed a voice only (low-process control) in setting budgets experienced significantly greater budget and task satisfaction than those allowed no input. Further, we also find, as did Lindquist, that participants with a high process control (vote for standard only) are significantly less satisfied with the task and budgets received than those with a low form of process control (voice only), thus supporting Lindquist's contention that vote is perceived as a form of pseudo-participation. Finally, unlike Lindquist (1995), our study finds a significant positive relationship between voice and performance in the experimental task. Specifically, we find participants who received a voice perform better—that is, they made more quality paper quilts—than others who did not receive a voice. Further, participants with a voice only outperform those with no input and, most importantly, also outperform participants with a vote only.

We believe our findings add not only to the management accounting literature but broadly to experimental research in the social sciences. That we reproduce the findings of Lindquist (1995) regarding voice and vote speaks to the importance of following the methods of the original research as closely as possible in a replication. Only then can practitioners extend findings and make decisive conclusions such as the problem of too much participation in standard setting backfiring for a manufacturer when workers must accept unattainable stretch targets. Sometimes it is better to just give employees a chance to air their grievances and leave it at that. We also find that workers who talk out their feelings about an unpleasant outcome, that is, high stretch performance standards, are significantly more productive than those with no chance for input.

Methodologically, we believe our findings impact experimental research across all the social sciences. Experimental tasks, measured in controlled laboratory studies, aim to emulate real-world occupations. In management accounting, these tasks are often meant to represent a production line. An experimental manufacturing task should be repetitive and boring, forcing participants to focus on the task at hand and the piece-rate payoff. Most research extending Lindquist (1995), however, employs

mental-based tasks and unsurprisingly finds diverse and sometimes opposing outcomes to Lindquist's. We develop and employ a new manual-based experimental task and match Lindquist's findings exactly. Interestingly, we also find a positive performance effect for voice and performance. Perhaps the learning effect in our study when building a six-piece paper quilt as compared to a 26-piece toy castle in Lindquist (1995) is much smaller. In that sense, participants producing paper quilts might better represent production workers who already know their jobs. If so, paper quilt making is likely a more appropriate manual-based task than castle making for experimental studies.

A potential limitation of our research is that the replication is conducted in Europe, opening the possibility of differences due to culture between our population and Lindquist's (1995) U.S. subject pool. The preponderance of national culture research, however, suggests that residents of Western and Central Europe share traits with those of the United States (e.g., Congden et al. 2009). It is possible also that experimenter bias could be driving our results. We did, however, make every effort to avoid predisposition in our findings by assigning experimenters to conduct two of the four conditions. We chose two conditions, as becoming an expert in all four would have required more than 1.5 h of training. We also didn't want experimenters to be confused as to which condition they were executing. In our study, we also assume that participants are motivated to earn a piece-rate incentive and that the truth-inducing incentive scheme motivates them to obtain a standard at a maximum attainable level to maximize their income. It is conceivable, however, that participants' self-noted attainable ranges of production ability may not be purely a measure of ability but also of motivation. In that sense, participants might be yoked (matched) on not only ability but motivation as well. Finally, as in all laboratory experiments, researchers sacrifice some external validity for internal validity in the controlled environment.

A logical and necessary extension of our research would directly test manual-versus mental-based tasks. Specifically, in the organizational justice context of this paper, researchers could measure the impact of voice and vote on satisfaction and performance using paper quilt making versus symbol decoding. That study would also likely include measures of motivation (intrinsic versus extrinsic). A study measuring differences in outcomes between manual and mental tasks could be conducted in any of the social sciences and add value to methods research. Another extension of our study could involve using the same manual task to measure the impact of various experimental operationalizations of voice and vote on satisfaction and performance. A study such as this could advance understanding of the impacts on satisfaction and performance of different experimental metrics of procedural justice. It could also serve as a methods analysis regarding the efficacy of making inferences about theoretical constructs when they are operationalized in multiple ways. It would also be interesting to experimentally test learning differences, as they relate to the amount of time it takes to learn new manual versus mental tasks. This contrast could also include similar measurements among various types of tasks.

In summary, a field of study has emerged from Lindquist's (1995) introduction of concepts of justice to the accounting literature. Subsequent research has examined many additional moderators and mediators of the impact of justice on satisfaction

and performance. Our findings suggest researchers should reflect on the soundness of their methods as they pursue future replications or extensions of seminal works to ensure they have captured the same underlying concepts. Only then can confident conclusions and inferences result.

Supplementary Information The online version of this article (<https://doi.org/10.1007/s00187-021-00318-3>) contains supplementary material, which is available to authorized users.

Funding Open access funding provided by University of Klagenfurt. (PricewaterhouseCoopers, The University of Northern Iowa, Verein zur Förderung der Wirtschaftswissenschaften an der Alpen-Adria-Universität Klagenfurt).

Compliance with ethical standards

Conflicts of interest (not applicable).

Availability of data and material (data available from corresponding author).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review*, 3(3), 185–201.
- Birnberg, J., & Nath, R. (1968). Laboratory experimentation in accounting research. *The Accounting Review*, 43(1), 38–45.
- Block, J., & Kuckertz, A. (2018). Seven principles of effective replication studies: Strengthening the evidence base of management research. *Management Review Quarterly*, 68(4), 355–359.
- Brandt, M., & IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J., & van't Veer, A. (2014). The replication recipe: What makes for a convincing replication. *Journal of Experimental Social Psychology*, 50, 217–224.
- Byrne, S., & Damon, F. (2008). To participate or not to participate? Voice and explanation effects on performance in a multi-period budget setting. *The British Accounting Review*, 40(3), 207–227.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436.
- Carpenter, J., & Huet-Vaughn, E. (2019). Real-effort tasks. In A. Schram & A. Ule (Eds.), *Handbook of research methods and applications in experimental economics* (pp. 368–383). Cheltenham, UK, & Northampton, US: Edward Elgar Publishing.
- Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149, 74–87.
- Chong, K., & Strauss, R. (2017). Participative budgeting: The effects of budget emphasis, information asymmetry and procedural justice on slack - Additional evidence. *Asia-Pacific Management Accounting Journal*, 12(1), 181–220.
- Chow, C. (1983). The effects of job standard tightness and compensation scheme on performance: An exploration of linkages. *The Accounting Review*, 58(4), 667–685.

- Chow, C., Lindquist, T., & Wu, A. (2001). National culture and the implementation of high-stretch performance standards: An exploratory study. *Behavioral Research in Accounting*, 13(1), 85–109.
- Congden, S. W., Matveev, A. V., & Desplaces, D. E. (2009). Cross-cultural communication and multi-cultural team performance: A German and American comparison. *Journal of Comparative International Management*, 12(2), 73–89.
- Cropanzano, R., & Folger, R. (1991). Procedural justice and worker motivation. In R. M. Steers & L. W. Porter (Eds.), *Motivation and Work Behavior* (5th ed., pp. 131–143). New York: McGraw-Hill.
- Darley, W. K. (2000). Status of replication studies in marketing: A validation and extension. *Marketing Management Journal*, 10(2), 121–132.
- Daumoser, C., Hirsch, B., & Sohn, M. (2018). Honesty in budgeting: A review of morality and control aspects in the budgetary slack literature. *Journal of Management Control*, 29(2), 115–159.
- Derfuss, K. (2016). Reconsidering the participative budgeting-performance relation: A meta-analysis regarding the impact of level of analysis, sample selection, measurement, and industry influences. *The British Accounting Review*, 48(1), 17–37.
- Eckartz, K., Kirchkamp, O., & Schunk, D. (2012). *How do incentives affect creativity?* Jena: Jena Economic Research Papers.
- Fisher, C. D. (1993). Boredom at work: A neglected concept. *Human Relations*, 46(3), 395–417.
- Francis-Gladney, L., Welker, R., & Magner, N. (2008). An experimental study of the effect of budget favorability on the formation of pseudo participation perceptions. *Advances in Management Accounting*, 17, 217–245.
- Gomez-Ruiz, L., & Rodriguez-Rivero, E. (2018). The motivational role of consultative participation in a multi-period target setting: An experimental study. *Spanish Journal of Finance and Accounting*, 47(3), 329–351.
- Habran, Y., & Mouritsen, J. (2020). Making intensity of efforts the same: Commensuration work in target-setting practices. *European Accounting Review*. <https://doi.org/10.1080/09638180.2020.1832901>.
- Hensel, P. G. (2019). Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015. *European Management Journal*, 37(1), 45–57.
- Indriani, M. (2015). Equity theory in accounting research. *International Journal of Economics, Commerce and Management*, 3(5), 907–924.
- Keller, J., & Bless, H. (2008). Flow and regulatory compatibility: An experimental approach to the flow model of intrinsic motivation. *Personality and Social Psychology Bulletin*, 34(2), 196–209.
- Kelly, K., Webb, R. A., & Vance, T. (2015). The interactive effects of ex post goal adjustment and goal difficulty on performance. *Journal of Management Accounting Research*, 27(1), 1–25.
- Kepes, S., Bennet, A. A., & McDaniel, M. A. (2014). Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management*, 13(3), 446–466.
- Lau, C. M., & Lim, E. W. (2002a). The effects of procedural justice and evaluative styles on the relationship between budgetary participation and performance. *Advances in Accounting*, 19, 139–160.
- Lau, C. M., & Lim, E. W. (2002b). The intervening effects of participation on the relationship between procedural justice and managerial performance. *The British Accounting Review*, 34(1), 55–78.
- Lau, C. M., & Tan, S. L. (2005). The importance of procedural fairness in budgeting. *Advances in Accounting*, 21, 333–356.
- Lau, C. M., & Tan, S. L. (2006). The effects of procedural fairness and interpersonal trust on job tension in budgeting. *Management Accounting Research*, 17(2), 171–186.
- Libby, T. (1999). The influence of voice and explanation on performance in a participative budgeting setting. *Accounting, Organizations and Society*, 24(2), 125–137.
- Libby, T. (2001). Referent cognitions and budgetary fairness: A research note. *Journal of Management Accounting Research*, 13(1), 91–105.
- Libby, T. (2003). The effect of fairness in contracting on the creation of budgetary slack. *Advances in Accounting Behavioral Research*, 6(1), 145–169.
- Liessem, T., Schedlinsky, I., Schwering, A., & Sommer, F. (2015). Budgetary slack under budget-based incentive schemes: The behavioral impact of social preferences, organizational justice, and moral disengagement. *Journal of Management Control*, 26(1), 81–94.

- Lind, E. A., Kanfer, R., & Earley, P. C. (1990). Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments. *Journal of Personality and Social Psychology*, 59(5), 952–959.
- Lindquist, T. (1995). Fairness as an antecedent to participative budgeting: Examining the effects of distributive justice, procedural justice and referent cognitions on satisfaction and performance. *Journal of Management Accounting Research*, 7(1), 122–147.
- Loukidou, L., Loan-Clarke, J., & Daniels, K. (2009). Boredom in the workplace: More than monotonous tasks. *International Journal of Management Reviews*, 11(4), 381–405.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316.
- Morrison, R., Matuszek, T., & Self, D. (2010). Preparing a replication or update study in the business disciplines. *European Journal of Scientific Research*, 47(2), 278–287.
- Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., et al. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology*, 113(1), 34–58.
- Nahartyo, E. (2013). Budgetary participation and procedural justice: Evidence from stretch budget condition. *Global Journal of Business Research*, 7(4), 85–100.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Ortmann, A. (2017). Book review: Replication in experimental economics. *Journal of Economic Psychology*, 59, 176–179.
- Otley, D., & Pollanen, R. (2000). Budgetary criteria in performance evaluation: A critical appraisal using new evidence. *Accounting, Organizations and Society*, 25(4–5), 483–496.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press & San Antonio, TX: The Psychological Corporation.
- Safkaur, O., & Pangayow, B. (2020). The effect of organizational justice on the budgetary slack of local financial management. *Journal of Talent Development and Excellence*, 12(1), 3115–3130.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100.
- Shin, J., & Grant, A. M. (2019). Bored by interest: How intrinsic motivation in one task can reduce performance on other tasks. *Academy of Management Journal*, 62(2), 415–436.
- Skarlicki, D., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology*, 82(3), 434–443.
- Sudarwan, D. (2019). The influence of budgetary participation on managerial performance with distributive fairness, procedural fairness, and budget goal commitment as mediating variables. *Science, Engineering and Social Science Series*, 3(5), 182–188.
- Thibault, L., & Walker, L. (1975). *Procedural justice: A social psychological analysis*. Hillsdale, NJ: Lawrence Elbaum Associates.
- Tyler, T. R. (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of Personality and Social Psychology*, 67(5), 850–863.
- Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, 7(4), 349–361.
- Vroom, V., & Yetton, P. (1973). *Leadership and Decision-making*. Pittsburgh: University of Pittsburgh Press.
- Young, M. (1985). Participative budgeting: The effects of risk aversion and asymmetric information on budgetary slack. *Journal of Accounting Research*, 23(2), 829–842.
- Young, M., Fisher, J., & Lindquist, T. (1993). The effects of intergroup competition and intragroup cooperation on slack and output in a manufacturing setting. *The Accounting Review*, 68(3), 466–481.
- Zainuddin, S., & Isa, C. R. (2019). The role of workplace fairness and information sharing in a budget setting process: An empirical study. *Gadjah Mada International Journal of Business*, 21(2), 135–158.

Authors and Affiliations

Tim M. Lindquist¹ · Alexandra Rausch²

✉ Tim M. Lindquist
timothy.lindquist@uni.edu

¹ PricewaterhouseCoopers, Professor of Accounting, The University of Northern Iowa,
Cedar Falls, IA, USA

² Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria