LET THE EVIDENCE SPEAK AGAIN!

FINANCIAL INCENTIVES ARE MORE EFFECTIVE THAN WE THOUGHT

JASON D. SHAW
Chair Professor and Co-Director, Centre for Leadership and Innovation
The Hong Kong Polytechnic University

NINA GUPTA
Distinguished Professor and John H. Tyson Chair of Management
University of Arkansas

To appear in *Human Resource Management Journal*

June 3, 2015
LET THE EVIDENCE SPEAK AGAIN!
FINANCIAL INCENTIVES ARE MORE EFFECTIVE THAN WE THOUGHT

In an editorial nearly two decades ago (Gupta & Shaw, 1998), we outlined the evidence-based case for the relationship between financial incentives and individual performance. Drawing on the findings of a meta-analysis of 40 years of financial incentives research (Jenkins, Mitra, Gupta, & Shaw, 1998), as well as Cameron and Pierce’s (1994) meta-analysis of a broader array of rewards, we showed that financial incentives were indeed strongly and positively related to individual performance and further, that incentives appeared to have no negative bearing on intrinsic motivation, as a number of influential scientists (e.g., Deci & Ryan, 1985; Pfeffer, 1998) and other writers (e.g., Kohn, 1993) had suggested previously. The research-based conclusions at that time seemed convincing, if not conclusive.

We hoped that the cumulative evidence would change the tone of the conversation in the literature and popular press away for the notion that financial incentives are harmful and toward a comprehensive exploration of the when and why of the effectiveness of financial incentives. Instead, the contrarian voices appeared to became louder, more varied and, discouragingly, more popular. A new generation of speakers and authors emerged, covering much the same ground, but with renewed vigor and technological sophistication. With a layperson’s confidence in the tradition of Alfie Kohn (1993, 1998), Dan Pink is a case in point. His best-selling book (Pink, 2009) and ever-popular internet speeches (nearly 15 million views at the time of this writing) ignore the accumulated scientific evidence and lambaste pay-for-performance while purporting to unlock the mystery of human motivation. This is despite the
fact that there is much more scientific evidence now, including primary studies, qualitative reviews (Fang & Gerhart, 2012; Gerhart & Fang, 2014), and meta-analytic summations (Cerasoli, Nicklin, & Ford, 2014; Garbers & Konradt, 2014). Not surprisingly, the cumulative evidence shows that financial incentives relate positively to performance, do not reduce intrinsic motivation, and, in general, are more effective than we previously thought.

In this essay, we rejoin the debate. In the following paragraphs we (1) briefly review the state of the literature in 1998, (2) highlight new meta-analytic findings and update our conclusions regarding the financial incentives \( \rightarrow \) performance relationship, (3) address the myth that financial incentives erode intrinsic motivation, (4) provide explanations for the presumed failure of financial incentives, and (5) offer some concluding thoughts and suggestions for moving forward. In terms of caveats, our review focuses on individual performance outcomes, primarily performance in workplace settings, although it does include meta-analytic and other evidence from non-work samples.

We take a broad view of individual performance, but roughly categorize it into quantity (how much is produced) and quality (how well something is done) dimensions. We also view individual performance from the organization’s perspective, acknowledging that individuals and other constituents (labor unions) may differ in their view of the level of individual performance, but also in terms of the importance of certain performance dimensions (e.g., Guest, 2011; Thompson, 2011). We focus on the narrow case of incentives and individual performance, recognizing not only that incentives may not be the best solution in every situation, but also that factors such as work complexity, coordination requirements, performance measurement differences, and monitoring systems can affect the decision to use incentives as well as the type
of incentive offered (Marsden & Belfield, 2010). In addition, a number of reviews and editorials have criticized performance-oriented HRM research for a positive normative bent as well as for ambiguity in the causal sequence between HRM practices and performance (e.g., Kaufman, 2010; Shin & Konrad, in press). These essays are largely irrelevant for our analyses. This review focuses on individual-level research with stringent designs in which the use of financial incentives precedes individual performance measurements.

The 1990s – Evidence Shows that “Financial Incentives are Effective”

Prior to the mid-1990s, the organizational literature on the effectiveness of financial incentives was rather sparse, but was also lacking in systematic summary evaluations. The most thorough of the summaries came in the form of qualitative reviews such as Jenkins (1986), Lawler and Jenkins (1992), and Gerhart and Milkovich (1992). Jenkins (1986) compared financial incentive studies conducted in the laboratory, the field, and in simulations using a qualitative “voting method.” He concluded that the evidence indicated that financial incentives had positive effects on performance quantity, and that effect sizes were generally similar across research designs. The first quantitative summary of the extrinsic reward literature was conducted by Cameron and Pierce (1994). Their meta-analysis of 96 extrinsic rewards studies using between-group experimental designs with control groups showed, contrary to assertions by Deci and Ryan (1985), Lepper, Greene, and Nisbett (1973) and many others, that extrinsic rewards had no overall negative bearing on intrinsic motivation. The results of this study raised the ire of proponents of Cognitive Evaluation Theory (CET) in academia (Lepper, Keavney & Drake, 1996; Ryan & Deci, 1996) and on the speaking-circuit (Kohn, 1996). Cameron and Pierce (1996) responded by conducting robustness checks of their meta-analytic findings, which
yielded similar results (although Kohn’s [1996] suggestion to include a set of poorly-designed, no-control-group studies was not adopted).

The Cameron and Pierce (1994) study provided the first direct cumulative evidence that extrinsic rewards were effective and did not diminish intrinsic motivation. The studies were obtained from the psychology and education literature, however, and did not include investigations involving actual financial incentives, a legitimate point of criticism. We and others have highlighted substantive differences between paid work and other situations where individuals receive extrinsic rewards such as prizes or recognition; these differences could cloud or change the conclusions drawn from non-monetary incentive studies. Jenkins et al. (1998) attempted to rectify this situation by collecting and analyzing relevant financial incentives studies, published over a 40-year period. The inclusion criteria were stringent: eligible studies had to (a) use non-self-report performance measures, (b) include a control group or a premeasure of performance, and (c) use actual money as the incentive. Among the curious details of this meta-analysis was that for all we thought we knew about financial incentives and performance, the number of studies eligible for inclusion was modest – only 39, or about 1 study per year over four decades. The results were straightforward. Financial incentives showed a moderately strong positive relationship with individual performance quantity ($\rho = .34$), a relationship that was stronger in field settings ($\rho = .48$) and simulation studies ($\rho = .56$), than it was in laboratory studies ($\rho = .24$). Also of note in the results was that task type did not moderate the size of the incentives $\rightarrow$ performance relationship. An extension of Deci and Ryan’s (1985) CET perspective is that the effects of incentives should be more negative in situations where the tasks are intrinsically pleasing. The meta-analysis results showed not only
positive relationships across all task types, but that the magnitude of the positive association was the same in mundane and pleasing task situations.

Only six studies in Jenkins et al.’s (1998) analysis included measures of performance quality, or how well tasks were accomplished; financial incentives were not significantly related to performance quality, although the sign was positive ($\rho = .08$). The accurate, scientific conclusion from the results was that financial incentives are significantly and positively related to performance quantity, but not to performance quality. Much was made about this scientific conclusion; Kohn, for example, interpreted the performance quality findings, as somehow indicating that financial incentives are significantly and *negatively* related to intrinsic motivation (e.g., see Kohn, 1998). Such interpretations simply defy logic. But, the close of the 1990s left us with additional questions to be answered. The results showed clearly that financial incentives and performance were positively related, that declines in intrinsic motivation resulting from the extrinsic incentives were likely mythical, and that there was a dearth of scientific evidence on the relationship of financial incentives and performance quality.

*The New Millennium – Financial Incentives are More Effective than We Thought*

Although the meta-analytic evidence of the 1990s revealed a strong pattern of findings in favor of financial incentives, the paucity of empirical research opened up opportunities for additional primary research, both primary and summative. It is encouraging that the number of primary studies has slowly increased, both in the management literature and in other disciplines including economics, medicine, nursing, and education. Primary studies advance knowledge, but they also advance confusion, as proponents on both sides of the debate pick and choose studies which support their views. The need for summative studies has been
addressed recently in two new meta-analyses—Garbers and Konradt (2013) focused on the workplace and included studies involving monetary payments, and Carasoli et al. (2014) reported a broader meta-analysis of the effects of intrinsic motivation and extrinsic incentives in the school, work, psychological, and physical domains. These meta-analyses estimated relationships with incentives across a broader array of performance-related criteria and, in the latter case, provide strong evidence that extrinsic incentives and intrinsic motivation “are not necessarily antagonistic and are best considered simultaneously” (p. 980). In short, the conclusion from research in these intervening years is that financial incentives are more effective than we previously thought.

Similar to the inclusion criteria in Jenkins et al. (1998), studies in the Garbers and Konradt (2014) meta-analysis (1) were experimental or quasi-experimental designs with pretests or control groups, (2) focused on financial incentives (not including other extrinsic motivators), and (3) included non-self-report performance measures. Their overall meta-analytic finding ($\rho = .32$) for individual performance of all types was nearly identical to Jenkins et al.’s (1998) result, although Garbers and Konradt’s (2014) literature review yielded nearly four times more studies ($k = 116$). A primary extension of this new meta-analysis was the ability to address performance quality in addition to quantity, a key shortcoming in the prior studies. These authors discovered that the effect size for financial incentives and performance quality was actually larger ($\rho = .39$) than the estimate for performance quantity ($\rho = .28$). The association for mixed or undetermined performance outcomes (e.g., ratings of “overall” performance by supervisors) was also substantial ($\rho = .38$). Thus, although performance quantity studies remain the dominant type in the literature, the growing evidence suggests that
when carefully-designed financial incentives are used to predict performance quality, the relationship is just as strong, perhaps stronger, than the relationship with performance quantity.

Cerasoli et al. (2014) took a broader approach, surveying literature in the psychological, educational, medical, and business areas and focusing on research that specifically operationalized intrinsic motivation. Like Cameron and Pierce (1994), these authors took a broad view of extrinsic incentives including financial payments, prizes, and credits. Their findings were striking. When extrinsic incentives were not offered, the relationship between intrinsic motivation and performance was positive, modest, and significant ($\rho = .27$). When extrinsic incentives were offered, the relationship between intrinsic motivation and performance was stronger ($\rho = .45$ when extrinsic incentives were indirectly salient and $\rho = .30$ when extrinsic incentives were directly salient). The authors drew some interesting conclusions: “An unexpected main effect for incentivization was observed, such that the predictive validity of intrinsic motivation did not erode, but in fact increased in the presence of incentives. Thus, incentivization actually boosted the intrinsic motivation—performance link” (p. 996).

Two new meta-analyses from health-care (Giles, Robalino, McColl, Sniehotta, & Adams, 2014) and psychological (Byron & Khazanchi, 2012) researchers provide strong ancillary evidence as well. Giles et al.’s (2014) meta-analysis revealed that financial incentives relate to a number of functional health-related behavior behaviors including smoking cessation (short- and long-term) and vaccination compliance. In a meta-analysis of 34 experimental and 8 non-experimental studies, Byron and Khazanchi (2012) found that individuals receiving incentives contingent on creative performance were indeed more creative.
The Myth of Intrinsic Motivation Erosion

The claim that extrinsic rewards erode intrinsic motivation is popular and persistent among professional speakers (e.g., Kohn, 1993; 1998; Pink, 2009) and academics (e.g., Pfeffer, 1998) and is pervasive across many disciplines including management, accounting, and education (see Gerhart & Fang, 2014, for a review). As noted above, this claim has been debunked quantitatively, but there are other qualitative issues as well.

Academically, the claim is rooted in CET (Deci & Ryan, 1985), and largely based on the findings from “free-time” studies (Gerhart & Fang, 2014), where the behavior of children playing a game is observed when extrinsic rewards are given and when they are withdrawn. Compared to control conditions, children in extrinsic rewards conditions tend to play pleasing games for a shorter period of time following the removal of the incentive. This effect is interpreted as evidence that the use of extrinsic rewards diminishes the children’s inherent pleasure in the game.

Several authors have pointed to problems with this interpretation. Gerhart and Fang (2014) questioned the applicability of “free-time” research to the workplace and to financial incentives. Bartol and Locke (2000) argued that “what people do during the time they are not being paid is of no importance” (p. 108). In addition, there is compelling evidence (acknowledged by CET founders themselves) that adults and children differ in their ability to separate the “informational and controlling aspects of rewards” (Deci et al., 1999: 656). Indeed, in workplace settings, evidence suggests that people feel more autonomy when they are paid for performance (Fang & Gerhart, 2012). These and similar issues led Gagne and Deci (2005) to acknowledge the inherent risks in using CET principles as a foundation for work motivation.
An additional concern not previously identified is that people could have negative reactions to the withdrawal of extrinsic rewards, not because money erodes intrinsic motivation, but because of the arbitrary fashion in which the incentive is removed—people do not like arbitrary and unexplained decisions about their rewards. A recent field experiment (Bareket-Bojmel, Hochman, & Ariely, in press) speaks to this issue. In a field experiment, the authors reported a productivity increase of more than 5% when short-term incentives (cash payments or vouchers) or verbal rewards were offered. But when monetary incentives were dropped without explanation, productivity declined. The authors explained their findings with the traditional CET view, under the assumption that workers’ intrinsic motivation was damaged. This explanation is untenable for several reasons. One, we should keep in mind that productivity increased when incentives were offered. Two, the bonuses appear to have been discontinued without explanation. Workers most likely were confused and angered by the injustice of the decision, particularly since their productivity had increased. The withdrawal of the incentive could well symbolize a punishment for higher productivity. In other words, the problem resides in the way the incentives were administered (and arbitrarily deactivated) rather than in the use of extrinsic incentives per se. A surfeit of studies on pay fairness (Shaw & Gupta, 2001; Shaw, in press), unexplained variation in pay (Conroy, Gupta, Shaw, & Park, 2014; Kepes, Delery, & Gupta, 2009; Shaw, Gupta, & Delery, 2002; Shaw, 2014; Trevor, Reilly, & Gerhart,, 2012), explanations and procedural justice (e.g., Shaw, Wild, & Colquitt, 2003) tell us clearly that individuals will react negatively – attitudinally and behaviorally – when unfair and fickle decisions are made about financial incentives and other important workplace issues.
In short, the corrosive effects of financial incentives on intrinsic motivation in the workplace are mythical. The evidence conclusively demonstrates their beneficial effects instead (Cerasoli et al., 2014). What the CET and Bareket-Bojmel et al. (in press) studies show are the negative effects of treating people arbitrarily and unjustly, a point long emphasized by justice and compensation researchers.

**But What About the Disconfirming Evidence?**

We noted earlier that researchers and professional speakers often pick and choose among studies to support their particular arguments. The scientific and popular literatures include examples of the failures of financial incentives—studies showing weak or negative effects of financial incentives on individual performance and other outcomes. This raises the question: If financial incentives effective, why do disconfirming studies appear in the literature? There are at least two responses to this conundrum. For one, scientists understand that the results of a given study may not be representative of the nature of a relationship in the population of studies. There are differences across studies in terms of sample characteristics, sample size, research design, measurement, and, of course, error, which could lead to somewhat different findings across primary studies. As an example, Park and Shaw’s (2013) meta-analysis of the relationship between turnover rates and organizational performance revealed a stable, moderate-in-magnitude, literature-level correlation between these variables. But, a stem-and-leaf plot of the correlation distribution across studies reveals a number of studies with near-zero association and a few with positive associations. It would be simple enough to cherry-pick a set of studies and make a case that organizations should strive to increase turnover rates in order to improve performance. This would, of course, be misleading,
inaccurate, and irresponsible. This is why meta-analytic results are important. They summate across the peculiarities and idiosyncrasies of individual studies and give us an indication of the sign and magnitude of the association, in general. In the case of incentives, summative evidence clearly shows financial incentives relate positively to performance quality and quantity and not only have no negative bearing on intrinsic motivation, but seem to enhance its potency in predicting performance.

Two, a closer examination of many of the “failures” reveals design and other problems which were most likely responsible, as in the case of the Bareket-Bojmel et al. (in press) study discussed above. Almost forty years ago, Lawler and Rhode (1976) and Kerr (1975) cautioned against the dysfunctional consequences of control systems and the folly of rewarding A while hoping for B. Unfortunately, many of these cautions are ignored in the research on financial incentives. When these issues are accounted for, we find that financial incentives often work exactly as they are designed, although not always as the designers hoped. The underlying reasons for the supposed failures can be grouped broadly into three categories – problems with study design, problems with incentive system design, and problems with incentive system implementation. It is to these we turn next.

**Problems with Study Design.** Clear examples of problems with study design appear in the pay dispersion literature. Pay dispersion—the spread of pay across employees—is, at least in part, attributable to the use of financial incentives such as merit pay. Pay dispersion is sometimes reported as having negative effects on employee outcomes (Bloom, 1999; Pfeffer & Davis-Blake, 1992; Pfeffer & Langton, 1993). But, as Gerhart and Rynes (2003) noted, studies reporting such negative effects usually control for employee performance. Thus, the observed
negative effects of dispersion are attributable to non-performance (and not performance) factors. In other words, they capture non-financial-incentives-based variance in the spread of pay. When organizations spread pay based on legitimate performance-related criteria pay dispersion is positively related to performance (Kepes et al., 2009; Shaw et al., 2002; Trevor et al., 2012) and organizations are more likely to keep their best performers (Shaw, 2015). For more detailed reviews of problems in the design of pay dispersion studies, see Conroy et al. (2014), Gupta et al. (2012), and Shaw (2014).

Another example of problems in study design occurs in the research by Ariely and colleagues (e.g., Ariely, Gneezy, Lowenstein, & Mazar, 2009). This work is often cited by professional speakers to support the failure of financial incentives. But as Gupta and Conroy (2013) indicate, the studies focused on the size of the financial incentive and failed to include a no-incentive control condition. Although the results can be interpreted as demonstrating the effects of incentives of different size, they cannot be interpreted as indicating whether or not the use of financial incentives, in general, is effective.

Other problems in study design can also be identified. We offer these two simply as illustrative of cautions in extrapolating from the results of previous studies.

**Problems with Incentive System Design.** Lawler and Rhode (1976) proposed that, for a control system to be most effective, it must be complete, objective, and influenceable. Even a cursory glance at the popular reports of ineffective incentive systems demonstrates a clear lack of attention to these recommendations in incentive system design. Since most research on financial incentives explicitly or implicitly assumes performance-based incentives (although
incentives necessarily needn’t be limited to performance), we focus on performance-related issues below, and address each of Lawler and Rhode’s recommendations in turn.

That *incomplete* measures lead to significant problems has been well-documented in the popular literature. Incomplete measures occur partly because of what Kerr (1975) describes as a fascination with the objective criterion. For many jobs, objective measures are simply not available for all critical aspects of the job, and organizations end up focusing on the partial set for which objective measures are available. Employees then attend only to the measured aspects, ignoring the remaining significant job duties. For example, in the Atlanta Public School cheating scandal (2015; Winerip, 2013), the incentive system focused exclusively on standardized test scores—teachers and administrators in the school district took legitimate and illegitimate steps to increase standardized test scores. In another example, the Pacific Gas and Electric Company (PG&E) offered bonuses to supervisors whose crews found fewer leaks and kept repair costs down (SFGate, 2011). An investigation concluded that bonuses for finding fewer leaks encouraged crews to ignore safety threats. In 2010, a blast killed eight people and destroyed 38 homes. Unfortunately, the incentive system worked—the crews found fewer leaks! Had the performance measurements been more complete and more accurate, the disaster could have been averted. But our favorite example comes from bus drivers in an Asian city (Govindarajan & Srinivas, 2013). Drivers were paid bonuses based on whether they reached their destination on time. One of the authors waited at the bus stop during peak traffic time. Several buses drove by without stopping, even though there were clearly empty seats. In other words, the incomplete incentive system encouraged drivers not to pick up passengers during rush hour, when the most fares (revenue for the organization) would be paid. These are but a
few examples of incentive systems working too well, promoting exactly the behaviors that are rewarded. Again, the problem is in design, not in the use of money.

As noted, Lawler and Rhode (1976) encouraged the use of objective (i.e., not based on subjective judgments) measures, but objective measures are typically not available for all aspects of performance, i.e., objective measures tend to be incomplete and suffer from criterion deficiency. Often, because complete objective measures are difficult to obtain easily, we settle for proxies that are easily at hand. But careful thought could produce an almost complete list of objective measures. For example, Arthur and Aiman-Smith (2001) reported that a joint union-management team developed a customized gainsharing formula (or family of measures) covering essential aspects of performance. This formula consisted of ten different dimensions that could be benchmarked historically. The lesson is that objective measures need not necessarily suffer from criterion deficiency, but that more effort is needed to develop a network of objective measures that encompass performance fully.

Still, the fact remains that complete objective measures may not always be available. As an alternative, organizations often use subjective performance appraisals by supervisors. Such merit systems, based on supervisory performance ratings, are arguably the most common basis for financial incentives (Heneman & Werner, 2005). But supervisory ratings of performance suffer from criterion contamination, and represent many other influences beyond employee performance (Murphy, 2008). Indeed Hoffman, Lance, Bynum, and Gentry (2010) demonstrated that idiosyncratic rater source effects accounted for 55% of variance in multisource performance appraisals, whereas performance dimensions accounted for 10% of the variance. These estimates are consistent with the results obtained by Scullen, Mount, and Goff (2000).
and Mount, Judge, Scullen, Systma, and Hezlett (1998). Performance ratings may overcome the problem of incomplete measures, but they introduce a variety of other problems. Unless the performance appraisal system is thoroughly examined for validity, tying financial incentives to them is likely to lead to undesired behaviors such as impression-management and politics.

A good example of the problems inherent in the use of incentive systems with uninfluenceable measures can be seen in recent scandal involving the Veterans Health Administration Scandal of 2014 (2015). Rewards in this case were tied to wait times for patient appointments, but no additional resources were provided to make shortening wait times feasible. Furthermore, because of the Iraq and Afghanistan wars, the caseload increased dramatically. Because already overworked workers had no legitimate way to influence the measure (wait times), they resorted to fudged data. Again, the incentive system worked, but bad design (uninfluenceable performance outcomes) led to substantial problems.

Beyond these criterion issues, at least two other problems with the design of incentive systems can be highlighted. One concerns the size of incentives, and the other the simultaneous use of multiple incentives.

In terms of incentive size, research shows that performance-based pay raises must be large enough (about 7%) to have meaningful effects (Mitra, Jenkins, and Gupta, 1997, Mitra, Tenhiäla, & Shaw, in press). Often, pay raises do not meet this threshold and are thus unlikely to have the desired effects. At the high end, we need more research on the sensitivity of performance to financial incentives levels. For many years, we have been proponents of research designed to understand when the positive performance effects of larger incentives begin to diminish (e.g., Shaw, 2011). The evidence that does exist, however, is not particularly
applicable to real-world incentive contexts. For example, in Ariely and colleagues’ (e.g., Ariely et al., 2009) research, individuals in a rural village in India were invited to play some games in return for cash prizes, the largest amounting to half of the typical annual wage in the village. A second study was conducted among only 24 undergraduate students and also included an exorbitant reward condition. These authors found that extremely large incentives were associated with some detrimental outcomes—a finding the authors attributed to choking. These findings are not surprising; surely, it is possible to induce anxiety, choking, cognitive inference and the like with such large prizes. What is needed are more reasonably designed studies that give us some indication of when positive returns to incentives begin to attenuate. In research terms, more evidence on the non-linearity of performance effects within reasonable incentive ranges is needed. In practical terms, prudence should be used and care taken in designing reasonable reward systems.

Sometimes organizations use more than one incentive system, and these incentives work at cross purposes. Beer and Cannon (2004), for example, reported on the problems encountered in the implantation of pay-for-performance plans in multiple sites at Hewlett-Packard in the 1990s. Each site developed its own plan, and their case reports demonstrate many design problems, but we focus on the use of multiple systems here. At the San Diego site, the system was designed to have a team pay-for-performance component and an individual skill-based pay component that encouraged learning new skills. The incentives systems were dropped after one year in one division, and in the rest thereafter. The training necessitated by skill-based pay plans implies that, for at least a modicum of time, employees will have sub-par performance. Implementing a team-performance-based pay plan concurrently militates against
employees in training mode. Indeed, many employees were resentful of the need for training, and ultimately the team pay-for-performance plan overwhelmed the skill-based pay plan. In other words, the plan failed not due to the use of team incentives (they work, in general; see, Garbers & Konradt, 2013) or because of the use of skill-based pay plans (they are shown to work in other settings; see, Shaw, Gupta, Mitra, & Ledford, 2005), but because the two were used simultaneously, and the two promoted conflicting behaviors.

**Problems with Incentive System Implementation.** Even when an incentive system is well-designed, problems can occur in implementation. Two issues are particularly salient in this regard. We addressed the first issue above—that supervisors and managers take arbitrary, unjust, and unexplained actions that underlie the negative effects observed in free-time or post-reward studies. Many years ago, Lawler (1978) noted that the implementation of innovative financial incentives requires the full commitment of middle as well as top management. This is in part because no change can be implemented perfectly, and that “kinks” are likely to develop during implementation. How managers address these “kinks” is vital to the success or failure of the incentive system. Arbitrary choices inconsistent with the inherent philosophy of the incentive system are likely to cause problems. Shaw et al. (2005) found, for example, that employee involvement in the design of the system and supervisor support for the system were consistent predictors of system success (productivity and workforce flexibility) and plan survival eight years later.

A related issue that is often observed is that system designers underestimate the power of the incentive system (perhaps because they listened to Kohn and Pink!). They set low performance standards that employees meet easily, resulting in higher employee earnings than
anticipated. An example of this phenomenon was observed in the Beer and Cannon (2004) report on Hewlett-Packard. In the first six months, the employees were excited by the team pay-for-performance plan and outperformed expectations, and the payout was higher than expected. Instead of building on this success, managers punished employees for their productivity by adjusting standards. Needless to say, this action triggered significant resistance among employees, and was partly responsible for the eventual termination of the plan. These kinds of managerial reactions also explain the development of rate restriction norms in teams and promotes distrust of management.

Other examples can be observed. But our point is that, quite often, incentives work as they are designed to do. Standardized test scores were improved, the buses ran on time, there were fewer leak reports, etc. Individuals performed the behaviors that were rewarded. But these “successes” are used to demonstrate the failure of incentives. Boxall (2014) considered contingent compensation or performance-based pay to be “one of the most obviously dangerous ideas” (p. 587), citing the 2008-2009 global financial crisis as an example of the pernicious “power of ill-conceived approaches to HRM to do more harm than good” (p. 587). These statements illustrate vividly the common misattribution—the blame for bad design and implementation is laid at the door of financial incentives per se. Defective design and implementation are the true culprits, not the use of money. This is not to say that financial incentives work in all situations. Rather, this is to say that reflexive laying the blame on financial incentives precludes a systematic examination of the actual causes of failure. It prevents us from learning from our mistakes.
Concluding Thoughts

As with debates about whether the sun goes around the earth and whether there is climate change, the scientific evidence has spoken about financial incentives in work settings—they are effective, they improve performance quantity, they improve performance quality, and they do not erode, but rather enhance the potency of, intrinsic motivation. It is time to put the issue of whether they work to rest; it is time to attend to issues of how and why they work. It is time to consider the conditions under which, and the people for whom, they work best. Gupta and Shaw (2014) highlighted many fruitful areas for compensation research. We urge the academy of scholars to move away from settled debates toward these promising areas.

Some additional thoughts are in order. First, by arguing that financial incentives are effective, we are not arguing that only financial incentives are effective or that people are motivated by money alone. Clearly, we are motivated by challenge, engagement, autonomy, mastery, and other factors. Instead, we are arguing that, if we accept that financial incentives work, we can move more easily to questions such as the best combinations of intrinsic and extrinsic rewards, the best combinations of social and extrinsic rewards, the best combinations of intrinsic, social, and extrinsic rewards, and so forth. Second, our analysis makes a general assumption that pay-for-performance is normatively accepted within the context. Much has been made about whether individuals in different cultural contexts, across different occupations, and across different levels of the societal strata, are interested in and will be motivated by the proper use of performance-based pay (Shaw, 2014). For example, we highlighted poor incentive design as one reason why bus drivers in an Asian city passed up the opportunity to pick up riders during rush hour. It is possible that other factors were also in play,
such as the cultural appropriateness of the system as well as tendencies toward solidarity or cohesion among workers which may have contributed to these reactions (Marsden & Belfield, 2010). These are important issues to address. We need to make space to begin addressing them.

Kohn is fond of saying things such as “no controlled scientific study has ever found a long-term enhancement of the quality of work as a result of any reward system” (1998; p. 34; see also, Kohn, 2009). We agree with him that we need longitudinal studies of incentives. This is a legitimate issue; such studies are scarce and investigations assessing the immediate and lasting effects of financial incentives are desirable (feasibility is another matter). One such study has recently appeared in the literature. Nyberg, Pieper, and Trevor (in press) studied the relationship between incentives and concurrent and future performance among nearly 12,000 employees over a 5-year period. Not surprisingly, given the meta-analytic evidence outlined above, they found that “merit and bonus pay, as well as their multiyear trends, as positively associated with future employee performance” (p. 1). But, we would point out here that existing theory does not suggest that financial incentives can be used as a behavior modification or a psychological panacea, inducing permanent changes in employee behaviors long after the incentives are in effect. Withdrawal of incentives may result in reversion to the status quo ante, i.e., financial incentives are effective only as long as they are in effect. Many of the studies highlighted by Kohn (2009) and by others involve assessments of the outcome in time periods where the incentives are no longer offered. It is not reasonable to expect that receipt of a financial incentive at one time period should relate to behaviors long in the future when incentives are no longer offered. Financial incentive decisions (and all other managerial decisions for that matter) are relevant only within their prescribed time windows. An illustrative
example here would be Yahoo CEO Marissa Mayer’s decision to end the long-standing work-from-home arrangement in the company (MacMillan, 2013). From media reports, we can reasonably conclude that employees had realized some benefits from flexible working hours in the past. But, it would be absurd to criticize “flex-time” as being unable to affect “long-term enhancement of the quality of work” after the program had been eliminated. Going forward, employees’ attitudes and performance would be best predicted by the working arrangements currently in place.

In conclusion, we encourage more research on critical compensation matters. We also pointed out many flaws in the design and implementation of incentive systems. Attention to these and similar issues is likely to result in better-designed incentive systems. Ultimately, we need to accept that financial incentives are effective, they are more effective than previously thought, and that therefore it is all the more important that they be used prudently.
References


Financial Incentives are Effective


