ソシオネットワーク戦略ディスカッションペーパーシリーズ ISSN 2434-9445
第 114 号　2024 年 10 月
RISS Discussion Paper Series
No.114　October, 2024

The unskilled-and-unaware problem and performance feedback in monotonous, easily accustomed, and repetitive work

Yasuhiro Nakamoto and Tomoharu Mori

文部科学大臣認定　共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

Research Institute for Socionetwork Strategies,
Kansai University
Joint Usage / Research Center, MEXT, Japan
Suita, Osaka, 564-8680, Japan
URL: https://www.kansai-u.ac.jp/riss/index.html
e-mail: riss@ml.kandai.jp
tel. 06-6368-1228
fax. 06-6330-3304

The unskilled-and-unaware problem and performance feedback in monotonous, easily accustomed, and repetitive work

Yasuhiro Nakamoto* and Tomoharu Mori

Abstract

We examine whether people predict their performance on monotonous, easily accustomed, and repetitive (MEAR) tasks accurately and confirm the effects of performance feedback on their MEAR work predictions. Considering typing on the keyboard as an MEAR task in modern society, we asked 128 university students to type a 12-digit number displayed on each monitor within 5 minutes, explaining that there would be 75 typing problems and participants would receive 20 yen per correct answer. Additionally, participants were informed that they could earn money by reaching a self-determined performance target, with the condition that if they did not reach their self-set goal, they would not receive any earnings. The main findings are as follows: Regardless of the performance levels of typing on keyboards, only 23% of the participants reached the goal, indicating that the absence of performance feedback led to individuals' overestimation of performance. Conversely, when participants received feedback on their own or/and others' performance information, at least 70% of them achieved their goals, significantly improving performance predictions in our task. Specifically, the effects between one's own and others' performance feedback did not vary at statistically significant levels.

Keywords: monotonous, easily accustomed, and repetitive work; unskilled-and-unaware problem; performance feedback

JEL Classification Code: C91, D12, D81, D84

1 Introduction

The unskilled-and-unaware problem (UUP) identified by Kruger and Dunning (1999), an influential study in this field, has shown that the unskilled lack the necessary metacognitive skills for accurate self-assessment predictions; hence, they overestimate their actual performance. Alternatively, the skilled more accurately predict theirs, an asymmetric error of performance predictions in metacognitive skill (see Dunning 2005; Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008 for reviews). Their argument for generating the UUP is that the unskilled have more difficulty recognizing their exact ability level than the skilled. Specifically, when the unskilled feel that a task is easy, they tend to strongly overestimate their performances (e.g., see Burson, Larrick, & Klayman, 2006). To reduce the overestimation of performance, specifically by the unskilled, numerous researchers (see Dunning, Heath, & Suls, 2004; Luckett & Eggleton, 1991; Nikiforow, 2019 for reviews) have focused on the role of performance feedback. Importantly, when examining the effects of performance feedback, some researchers (Arkes, Christensen, Lai, & Blumer, 1987; Burson et al., 2006, Studies 1 and 2; Lichtenstein & Fischhoff, 1980; Moore & Cain, 2007; Grossman and Owens, 2012; Ryvkin, Krajč, & Ortmann, 2012, Task 2 of Study 2; Stone & Opel, 2000) employed quizzes such as general knowledge questions. In other research (Hacker, Bol, Horgan, & Rakow, 2000; Ryvkin et al., 2012, Study 1; Sabater-Grande et al., 2023), undergraduate students who participated were asked to predict the scores of the examination in the class (see Table 1).

[Table 1 around here.]

Unlike the tasks in the abovementioned studies, we focus on monotonous, easily accustomed, and repetitive (MEAR) works, which are represented by sorting, packaging, inspection jobs in the manufacturing industry, data entry business in the service industry, and everyday activities such as cutting and washing while cooking and typing on the keyboard. Notably, MEAR works play an important role in creating employment opportunities and reducing unemployment, particularly for lower-skilled workers and people living in developing countries, which is an engine that supports economic growth. Thus, recognizing the skills of workers in MEAR work is important for opening up opportunities for advancement and leading to higher quality work. However, in this field, insufficient attention has been devoted to MEAR work. To our knowledge, our study is the first to focus on the UUP in MEAR work. Thus,

we examine whether people predict their MEAR work performance accurately. Based on the UUP, it can be expected that the unskilled cannot make accurate self-assessment predictions, while the skilled can. However, for MEAR tasks, there may not be a significant difference in performance between the unskilled and the skilled. If this is the case, the UUP will not be observed, and both groups will either be able to make accurate self-assessment predictions or both will be inaccurate. Furthermore, we confirm the effects of performance feedback on their predictions of MEAR work.

In this study, we regard typing on the keyboard as a typical MEAR task. This is because many people in modern societies, including our participants (undergraduate students in a Japanese university), are accustomed to typing on the keyboard and consider the work monotonous and repetitive. For instance, the institution in our experiment reported that almost all students had the opportunities to use the computers frequently. In detail, 76.2% of students had their own computer and 14.2% used the parent's computer in their own house. Moreover, the university has a lot of personal computers in each building, and the students can use them freely. We then conducted a laboratory experiment to test participants' performance predictions related to typing on the keyboard. Before performing the typing task, participants were asked to predict how well they could type. Thereafter, we tested whether the prediction was correct and what influenced the predictions.

There are two main differences between existing studies and ours. The first difference is the uncertainty of the next task. As quizzes and examination problems cover various kinds of content, the subsequent quiz and examination problems are unpredictable. Alternatively, the next task in MEAR work is predictable owing to the monotonous repetition. Therefore, people are able to predict their MEAR work performance more clearly compared to that in quizzes and exams. Moreover, as mentioned in Arkes et al. (1987), because a "trickery" hypothesis exists owing to the uncertainty of the quiz that comes up, people may be skeptical about whether they have been tricked. However, MEAR work involves little uncertainty and is not considered tricky. The second difference is the introduction of monetary incentives, which encourage people to make more accurate predictions (see Camerer and Hogarth (1999) for a review and Ryvkin, Krajč, and Ortmann (2012, Section 1.2)). In our study, participants were strongly motivated by monetary incentives to predict their performance more accurately. All participants applied for the experiment to obtain earnings, implying that their primary

concern was to earn money. Moreover, all participants were asked to set a performance target of typing on the keyboard in the real-effort task before implementing it. They were informed that they would obtain the targeted earnings on attaining the performance target; otherwise, they would receive nothing. In other words, participants wanted to increase their targeted earnings by setting higher goals; however, they recognized that it would increase the possibility of no earnings due to the non-achievement of self-set goals. Conversely, the monetary incentives underlying accurate predictions in the existing studies may be weak (See Table 1 for the reviews of related literature). Arkes et al. (1987), Burson et al. (2006), Lichtenstein and Fischhoff (1988), and Hacker et al. (2000) did not offer any incentive to provide accurate performance estimates. Although participants in studies by Stone and Open (2000), Ryvkin, Krajč, and Ortmann (2012), and Sabater-Grande et al. (2023) were motivated by monetary incentives, they were students enrolled in the class. Therefore, the undergraduate students' primary purpose was not to obtain earnings but to acquire education.

Considering MEAR work with monetary incentives, participants, who were randomly assigned to one of four groups, performed the common task of typing the 12-digit number displayed on each monitor within 5 minutes, such as "392832944950" and "958375930201." The four groups, called *Control*, *Exp*(erience), *Info*(mation), and *Info*(mation)*Exp*(erience), differed from each other in how the performance feedback was introduced. Participants in the Exp group took part in a practice task before the real-effort task, wherein they typed on the keyboard within 5 minutes and confirmed their performances in the practice task. In other words, they received feedback on their performances. Subsequently, participants in the Info group confirmed the distribution of others' performances collected prior to the investigation. Participants in the InfoExp group took part in the practice task and received the distribution of others' performance. Finally, participants in the Control group did not receive performance feedback in our experiment. See the detailed description in Section 2.

Two main findings emerged. First, although MEAR work performance is more predictable than that in quiz and examination problems, and the overestimation did not yield money, many participants overestimated their performances without feedback. Notably, not only poor but also high performers overestimated their performances, that is, there was no asymmetric prediction between the unskilled and skilled and no UUP was identified. Second, the performance feedback, irrespective of its kind, substantially reduced overestimation in MEAR

work. Specifically, considering the results in the Exp and Info groups, no statistically significant differences in effects were found between one's own and others' performance feedback. These findings are of relevance for a realistic society. For instance, when employers interview potential employees for MEAR work, the latter tend to overestimate their performances in the absence of performance feedback, regardless of actual performance levels. Alternatively, even if potential employees have not practiced beforehand, implying that employers can make budget and time cuts, the information on others' performance allows potential employees to predict their MEAR work performances accurately, which would make recruitment interviews more efficient.

Our research is also related to a series of studies on belief updating. These studies discuss whether updates occur when information is obtained and whether these updates are based on Bayes' rule. In a survey conducted by Santos-Pinto and de la Rosa (2020), it is demonstrated that overconfidence persists even with repeated feedback (e.g., Grossman and Owens, 2012). Our study shows that strong monetary incentives and information provision eliminate overconfidence, making a valuable contribution to the existing literature.

The remainder of this paper is structured as follows: Section 2 outlines the experimental design. Section 3 shows our findings. Finally, Section 4 presents the conclusion.

2 Materials and Methods

Our experiment participants, recruited from the pool of subjects maintained by the Center for Experimental Economics at Kansai University in Japan, were undergraduate students. There were 128 participants, including 59 men and 69 women (each session consisted of approximately 15 participants), and each participated only once. Our experiments were conducted in December 2019 at the experimental economics laboratory in the university. Each session lasted 30 minutes on average.

Upon arrival, participants were seated in isolated cubicles. We read the instructions aloud, explaining that participants would obtain a 1,000-Japanese-yen show-up fee (approximately 10 US dollars). Each subject signed a consent form for the abovementioned show-up fee, experiment time, and so on. Then, all participants were informed that their task would be to type the 12-digit number displayed on each monitor within 5 minutes, referred to as the

typing task in this study. Specifically, as mentioned by Giannouli (2011, 2013), performance in MEAR work for extended time may be influenced by perseveration. Thus, in our experiment, to prevent perseveration from complicating self-assessment predictions, we limited the typing time on the keyboard to a short duration of 5 minutes.

They confirmed samples of the randomly arranged 12-digit number on each monitor, such as "392832944950" and "958375930201." In addition, to limit the requisite skills further, we explained that they could select a familiar typing tool: the QWERTY keyboard or the numeric keypad (10 keys) next to the QWERTY keyboard. Next, we explained that there would be a total of 75 typing problems and that participants would receive 20 yen per correct answer; therefore, the maximum earnings would be 1,500 yen (=20 yen $\times$ 75 problems). In addition, we explained that they could not return to the previous problem after moving on to the following problem. All participants were informed about the role of goal setting, whereby they could obtain earnings by reaching a self-determined performance target. We explained that if the number of correct answers exceeded the self-set goal of correct answers, they could obtain "20 yen $\times$ the number of the self-set goal." In contrast, if participants did not reach the self-set goal, they would not receive any earnings. Thus, participants were motivated to set higher goals; however, higher goals could decrease the probability of reaching those goals. Furthermore, we gave a concrete explanation using an example as follows: Suppose that a participant sets 20 as the self-set goal; the obtained earnings would be 400 yen (=20 yen $\times$ 20 problems) if they achieved the self-set goal. Notably, even if the number of correct answers exceed 20, such as 25 and 75, the earnings do not change and amount to 400 yen based on the self-set goal. Thereafter, they were asked to reveal the targeted number of correct answers in the typing task. After completing the real-effort task, participants were asked to answer the post-experimental questionnaire.

To determine the sample size for this study, we conducted a power analysis. We aimed to detect a large effect size (Cohen's $d = 0.8$) with a significance level (α) of 0.05 and a desired power (1-β) of 0.8 for pairwise comparisons. Based on these parameters, the power analysis indicated that a minimum of 26 participants per group was required to achieve sufficient statistical power. To account for potential dropout, we slightly increased the sample size. Then, we created four groups called the Control ($n = 31$ of the total 128 participants), Exp ($n = 34$), Info ($n = 30$), and InfoExp ($n = 33$) groups, as shown in Table 2, where the

variables except for Age and Male are explained later.

[Table 2 around here.]

Each group was constructed as follows: First, participants in the Control group did not receive any performance feedback. Therefore, they did not have any way to recognize their performance. Second, in the Exp group, participants were additionally informed that the practice task would be implemented before the real-effort task, and they did not receive any earnings for the practice task. Participants in the Exp group then typed within 5 minutes for the practice task and confirmed their performance (the number of correct answers) on the monitor. Consequently, participants in the Exp group were able to recognize their achieved typing performances in our experiment. Third, in the Info group, participants confirmed the distribution of correct answers in the keyboard typing task from the other 30 participants before the real-effort task (see Figure 1). These 30 participants had been newly recruited in the branch school of the same university. To guarantee the reliability of Figure 1, following Luckett and Eggleton (1991), participants in the Info group were additionally informed about source attributes (i.e., the 30 participants in the prior investigation were the undergraduate students in another branch of the same university). By looking at Figure 1, they understood the mean, maximum, and minimum levels in addition to the distribution of the correct answer. For example, based on Figure 1, they might have understood that it would be difficult to achieve correct answers for all 75 problems within 5 minutes; alternatively, they might have recognized that they would be able to answer at least a minimum of 15 problems. Primarily, participants in the Exp group did not confirm others' performances in Figure 1, whereas those in the Info group did not practice beforehand; hence, they did not recognize their own performance. Alternatively, the InfoExp group was a mix of the Info $\times$ Exp groups. Therefore, participants in the InfoExp group practiced typing, similar to the Exp group, and confirmed the distribution of the correct answer, as in the Info group, implying that they were able to recognize both others' and their own performances.

[Figure 1 around here.]

Finally, after completing the real-effort task, participants were asked to answer the post-experimental questionnaire by providing the following information: gender, job, math tastes,

personality traits, and degree of loss aversion. As for the job, participants were asked to answer whether they had a part-time job (= 1) or not (= 0), which is denoted by the variable "Job" in Table 2. For math tastes, participants were asked whether mathematics was one of their favorite subjects, corresponding to the variable "Math" in Table 2. This is because participants whose favorite subject is not mathematics may have disliked the 12 randomly arranged numbers; therefore, their performances may have been poorer than those whose favorite subject is mathematics.

Next, considering that Kruger and Dunning, as well as their colleagues, have argued that a lack of metacognition by the unskilled is the main element that leads to the overestimation of performances, participants were asked to answer the Big Five personality test. This is because there is robust evidence that metacognitive beliefs are significantly correlated with the Big Five personality traits (e.g., Marino et al., 2016). For instance, when assessing the contribution of personality to metacognition on problematic Facebook use among young adults, Marino et al. (2016) show that emotional stability influences metacognition (negative beliefs about thoughts and cognitive confidence). To measure personality traits (see Table 3), we used the Japanese version of the 10-Item Personality Inventory (Oshio, Abe, & Cutrone, 2012; Oshio, Abe, Cutrone, & Gosling, 2013), which was a measure of the Big Five (extroversion, agreeableness, conscientiousness, neuroticism, and openness in Table 3) personality dimensions (Gosling, Rentfrow, & Swann, 2003). We averaged the ratings of the two related items, generating the score of the personality traits in each dimension. For instance, the score of extroversion was calculated based on scores of Questions A and F in Table 2, where the score given in Question F was arranged in the opposite direction because the content in Question A was opposite to that in Question F. Each dimension was rated with one positively and one negatively keyed item, and each item was scored on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Using Table 4, participants were asked to reveal loss-averse propensities in a coin-toss game, informing them that they would not receive any earnings in this problem. It could be predicted that participants who disliked loss tended to set a lower number of correct answers as their own goal because they strongly sought to avoid losing all earnings. For instance, if participants selected B (A) in all problems, the degrees of loss aversion would be very large (very small). We treated the switching timing as the data in the empirical analysis in Section

3; however, these data did not impact the self-set goals at statistically significant levels. The mean level and standard deviation are given by "Loss propensity" in Table 2.

[Tables 3 and 4 around here.]

3 Results and Discussion

Table 5 presents a summary of our experimental results. The table shows the mean and standard deviation of the number of correct answers in each group (Answer), the goal set by participants (Goal), and the indicator variable, indicating whether the goal was reached in the number of correct answers (Success). Figure 2 shows box plots for Answer and Goal. Figure 3 presents the scatter plots for Answer and Goal, with 45-degree lines for each group. As the 45-degree line shows all points where the number of correct answers is equal to the self-set goal, the points to the upper left of the 45-degree line mean that the number of goals is greater than the performance; hence, the participants did not achieve the goal. Conversely, the points to the bottom right of the 45-degree line mean that the number of goals is less than the performance, implying that the participants achieved the goal. Table 6 shows the group differences for each variable. It also presents the $p-$values for the $t-$test for the difference of means, Wilcoxon rank-sum test, and the test of equal variances.

[Tables 5 and 6 around here.]

[Figure 2 and 3 around here.]

In the Control group, the mean Goal level is 40.48, the highest among all groups. Compared to the other groups, the $p-$values are less than 0.001 in the $t-$test and Wilcoxson rank-sum test. Alternatively, the mean Answer level is 25.81 in the Control group, the lowest among all groups where the $p-$value is low only for the difference with the InfoExp group. These statistics imply that many participants do not achieve their goal. Success shows that only 23% of the participants in the Control group reached the goal, the lowest percentage among all groups. Figure 3 also shows that Goal exceeded Answer for many participants. Compared to the other groups, the $p-$values are less than 0.001 in Fisher's exact test.

Although all participants received information about the task's properties, thereby reducing ambiguities regarding the task (Steinmann, 1976), only the Control group exhibited

extreme misjudgment regarding goal setting. Participants may have set extremely high goals, much above their performance levels, because of their overconfidence regarding typing on the keyboard arising from the following two reasons. First, existing studies (e.g., Brown, Farrington, & Sprinkle, 2016; Hoelzl & Rustichini, 2005; Moore & Cain, 2007) define an easy task as one that people generally feel capable of doing, with relatively high absolute performance, and argue that people tend to overestimate their performance due to overconfidence. As typing on the keyboard as MEAR work was easy for the undergraduate students who participated in our experiment, participants tended to overestimate their performances.

Second, only participants in the Control group did not have any recognizable feedback regarding task performance, unlike in the other groups. In other words, as those who participated in the Control group did not know about their own and others' performances, they had no choice but to judge their performances through their keyboard typing experience. However, Kahneman & Tversky (1979) mention that when making an intuitive prediction, such as goal setting in our study, people tend to neglect distributional information about the outcomes in similar cases, such as past experiences of keyboard typing, leading to a major error in intuitive prediction. If this were the case, participants in the Control group neglected only the clues for predicting their performance. Hence, the undergraduate students' ease of keyboard typing generated overconfidence about the typing task and led to an overestimation of expected performance.

In the Exp group, the mean value of Goal is 24.94, approximately 60% of that of the Control group. The standard deviation of Goal is reduced to less than half of that of the Control group. Tests of equal variances with the Control group indicate that the variances significantly differ. However, the mean value of Answer is 28.03, slightly higher than that of Goal. The value is higher than that for the Control group, but the difference is not significant. These results suggest that the success rate of the Exp group is as high as 76%, more than triple that of the Control group. These results indicate that experience increases the predictive power of one's productivity, in terms of both mean and variance, as well as the probability of success.

The result of the Info group is similar to that of the Exp group. The mean value of Goal is 24.30, approximately 60% of that of the Control group. This value is smaller than the Control group mean of 25.8, as shown in the experiment ($p = 0.088$ for the $t-$test). The

standard deviation of Goal is reduced to less than one-third of that of the Control group. The mean value of Answer is 26.50, almost equivalent to that of Goal, with little difference compared to the Control group. The success rate in the Info group is 70%. These values are close to their Exp group counterparts. For any difference, the p-value is higher than 0.1 and not significant. The results suggest that the same effect as in the practice task can be achieved by showing participants the distribution of correct answers in advance.

The results of the InfoExp group are highly different from those of the Control group and slightly different from the Exp and Info group results. The mean value of Goal is 27.70, significantly smaller than that of the Control group and slightly larger than the Exp and Info group values. The $p-$values for comparison between the Exp and Info group range between 0.01 and 0.1, suggestive of a difference. The combined effect of experience and information may have facilitated higher goals. The mean value of Answer is 32.03, the highest among all groups; the difference between the Control and Info groups is significant, suggesting a difference compared to the Exp group. The success rate is also the highest, at 88%, compared to the other groups. Figure 3 shows that Answer exceeds Goal for almost all participants in the InfoExp group and even those below the 45-degree line; the difference between Answer and Goal is small.

These results suggest that performance feedback is crucial for personal goal setting irrespective of the source of feedback (individuals or others) in our study. In addition, the effects of individuals' performance feedback or joint effects of individual and others' performance feedback generated improved performance.

We find that many participants overestimated their performances without any feedback information in the Control group. Interestingly, we did not see the UUP in the sense that almost all participants in the Control group overestimated their own performances irrespective of their typing skills. Intuitively, as participants in the Control group had to judge their performances through experience, the easy task of typing on the keyboard led to overconfidence and a large divergence between the self-set goal and actual performance. Next, the performance feedback greatly reduced the overestimation of MEAR work performance. Specifically, considering the results in the Exp and Info groups, as there was no statistically significant difference in the effects between one's own and others' performance feedback, performance feedback, regardless of the kind, would work efficiently.

Finally, Table 7 presents the results of the regression analysis of the variables analyzed so far on group dummies and individual characteristics. The odd columns denote models with only group dummies as independent variables, and the even columns represent models with individual characteristics added to the independent variables. The individual characteristics do not significantly influence the effect of group dummies in any of the analyses. The coefficients of the group dummies and adjusted coefficients of determination change slightly after adding the independent variables. Considering the individual coefficients, only Agreeableness and Neuroticism have a suggestive negative impact on goal setting. The negative coefficient of Agreeableness suggests that sympathetic, warm, and cooperative participants lowered their self-set goals to set goals similar to those of others. The negative coefficient of Neuroticism implies that anxiety and worry about whether goal achievement may adversely affect goal setting.

[Table 7 around here.]

4 Conclusion

Considering typing on the keyboard as a MEAR task in modern society, we re-investigated the effects of feedback on the predictions of one's own MEAR work performance. The main findings are presented as follows: First, irrespective of the actual performance levels, many participants overestimated their performance when no performance feedback was provided. Second, even if participants did not practice, the feedback regarding others' performances greatly improved their predictions, and its effect was the same as that of the feedback regarding their own performance.

The effects of feedback information on overestimation could potentially be applied to various tests and tasks beyond MEAR tasks, such as intelligence quotient (IQ) and emotional quotient (EQ) tests. Recent research (Giannouli, 2022) reports that younger males overestimate their IQ and EQ compared to younger females, while older females overestimate theirs compared to older males. Our study suggests that providing this research information to participants as preliminary information could significantly change the results.

Acknowledgments We thank the staff at the Center for Experimental Economics at Kansai

University, as well as Yoshio Kamijo and Kazuhito Ogawa.

Declaration of interest statement The authors hereby affirm that they have no financial nor material interest that relates to the research described in the paper "The unskilled-and-unaware problem and performance feedback in monotonous, easily accustomed, and repetitive work."

Funding This is a product of research that was financially supported (in part) by the Kansai University Fund for Supporting Young Scholars, 2018.

Ethical approval Ethical approval was obtained from the ethical review board of Kansai University (Approval Number: 2018023). The data obtained are treated confidentially and stored anonymously, which strictly complies with General Data Protection Regulations. Informed consent was obtained from all individual participants included in the study.

Accordance statement All methods were performed in accordance with the relevant guidelines and regulations.

Data accessibility statement The dataset in this study is available from the corresponding author upon reasonable request.

References

[1] Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. Two methods of reducing overconfidence. *Organ. Behav. Hum. Decis. Process.* 39(1), 133–144 (1987). https://doi.org/10.1016/0749-5978(87)90049-5

[2] Brown, J. L., Farrington, S., & Sprinkle, G. B. Biased self-assessments, feedback, and employees' compensation plan choices. *Account. Organ. Soc.* 54, 45–59 (2016). https://doi.org/10.1016/j.aos.2016.08.003

[3] Burson, K., Larrick, R. P., & Klayman, J. Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *J. Pers. Soc. Psychol.* 90(1), 60—77 (2006). https://doi.org/10.1037/0022-3514.90.1.60

[4] Camerer, C. F., & Hogarth, R. M. The effects of financial incentives in experiments: A review and capital-labor-production framework. *J. Risk Uncertainty* 19, 7—42 (1999). https://doi.org/10.1023/A:1007850605129

[5] Dunning, D. *Self-insight: Roadblocks and detours on the path to knowing thyself.* (New York: Psychology Press, 2005). https://psycnet.apa.org/doi/10.4324/9780203337998

[6] Dunning, D., Heath, C., & Suls, J.M. Flawed self-assessment. *Psychol. Sci. Publ. Int.* 5(3), 69—106 (2004). https://doi.org/10.1111/j.1529-1006.2004.00018.x

[7] Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. Why the unskilled are unaware: Future explorations of (absent) self-insight among the incompetent. *Organ. Bahav. Hum. Decis. Process.* 105(1), 98—121 (2008). https://doi.org/10.1016/j.obhdp.2007.05.002

[8] Giannouli, V. Music in a serial repetition task: is there perseverative behavior? *Acta Neuropsychologica* 9(4), 361—368 (2011). GICID:01.3001.0001.0791

[9] Giannouli, V. Number perseveration in healthy subjects: Does prolonged stimulus exposure influence performance on a serial addition task? *Adv. Cogn. Psychol.* 9(1), 15—19 (2013). https://doi.org/10.2478/v10053-008-0127-8

[10] Giannouli, V. Are sex differences in self-estimated intelligence an elusive phenomenon? Exploring the role of working memory, creativity, and other psychological correlates in young and older adults. *Brain Behav.*, 13(2) (2022). https://doi.org/10.1002/brb3.2857

[11] Grossman Z., & Owens, D. An unlucky feeling: Overconfidence and noisy feedback. *J. Econ. Behav. Organ.* 84 (2), 510—524 (2012). https://doi.org/10.1016/j.jebo.2012.08.006.

[12] Gosling, S. D, Rentfrow, P. J, & Swann, W. B. A very brief measure of the Big-Five personality domains. *J. Res. Pers.* 37(6), 502—528 (2003).
https://doi.org/10.1016/S0092-6566(03)00046-1

[13] Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. Test prediction and performance in a classroom context. *J. Educ. Psychol.* 92(1), 160—170 (2000).
https://doi.org/10.1037/0022-0663.92.1.160

[14] Hoelzl, E., & Rustichini, A. Overconfident: Do you put your money on it? *Econ. J.* 115(503), 305—318 (2005).
https://doi.org/10.1111/j.1468-0297.2005.00990.x

[15] Kahneman, D., & Tversky, A. On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cogn.* 7(4), 409—411 (1979).
https://doi.org/10.1016/0010-0277(79)90024-6

[16] Kruger, J., & Dunning, D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *J. Pers. Soc. Psychol.* 77(6), 1121—1134 (1999).
https://doi.org/10.1037/0022-3514.77.6.1121

[17] Lichtenstein, S., & Fischhoff, B. Training for calibration. *Organ. Behav. Hum. Perform.* 26(2), pp.149—171 (1980).
https://doi.org/10.1016/0030-5073(80)90052-5

[18] Luckett P. F., & Eggleton, I. R. Feedback and management accounting: A review of research into behavioural consequences. *Account. Organ. Soc.* 16(4), 371—394 (1991).
https://doi.org/10.1016/0361-3682(91)90028-D

[19] Marino, C., Vieno, A., Moss, A.C., Caselli, G., Nikčevic, A.V., & Spada, M.M. Personality, motives and metacognitions as predictors of problematic Facebook Use in university students. *Pers. Individ. Differ.* 101, 70—77 (2016).
https://doi.org/10.1016/j.paid.2016.05.053

[20] Moore, D. A., & Cain, D. M. Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organ. Behav. Hum. Decis.*

Process. 103(2), 197–213 (2007).
https://doi.org/10.1016/j.obhdp.2006.09.002

[21] Nikifolow, N. Behavioral effects of feedback in management accounting research: A review of the current state. (2019). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3563500

[22] Oshio, A., Abe, S., & Cutrone, P. Development, reliability, and validity of the Japanese version of Ten-Item Personality Inventory (TIPI-J). *Jpn. J. Pers.* 21(1), 40—52 (2012).

[23] Oshio, A., Abe, S., Cutrone, P., & Gosling, S.D. Big five content representation of the Japanese version of the Ten-Item Personality Inventory. *Psychol.* 4(12), 924—929 (2013).

[24] Ryvkin, D., Krajč, M., & Ortmann, A. Are the unskilled doomed to remain unaware? *J. Econ. Psychol.* 33(5), 1012—1031 (2012).
https://doi.org/10.1016/j.joep.2012.06.003

[25] Santos-Pinto, L., & de la Rosa, L. E. Overconfidence in labor markets, In: Zimmermann, K. (eds) Handbook of Labor, Human Resources and Population Economics. (Springer, Cham., 2020)
https://doi.org/10.1007/978-3-319-57365-6_117-1

[26] Sabater-Grande, G., Georgantzís, N., & Herranz-Zarzoso, N. Goals and guesses as reference points: A field experiment on student performance. *Theor. Decis.* 94, 249—274 (2023).
https://doi.org/10.1007/s11238-022-09892-x

[27] Steinmann, D. O. The effects of cognitive feedback and task complexity in multiple-cue probability learning. *Organ. Behav. Hum. Perform.* 15(2), 168—179 (1976).
https://doi.org/10.1016/0030-5073(76)90035-0

[28] Stone, E. R., & Opel, R. B. Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organ. Behav. Hum. Decis. Process.* 83(2), 282—309 (2000).
https://doi.org/10.1006/obhd.2000.2910

Table 1: Previous Studies

Study	Task	Monetary Incentives	Subjects	Feedback
Burson, Larrick, and Klayman (2006)				
Study 1	Quiz	Only show-up fee	University students	No
Study 2	Quiz	Only show-up fee	University students	No
Study 3	"Word prospector" game	Only show-up fee	University students	No
Arkes, Christensen, Lai, and Blumer (1987)				
Experiment 1	Quiz	No	17-33 aged subjects	Yes
Experiment 2	Quiz	No	17-37 aged subjects	Yes
Lichtenstein and Fischhoff (1980)				
Experiment 1	Quiz	Hourly payment	University students	Yes
Experiment 2	Quiz	Hourly payment	University students	Yes
Hacker, Bol, Horgan, and Rakow (2000)				
	Score in exams	No	University students	Yes
Stone and Opel (2000)				
	Quiz (art history)	Yes	University students	Yes
Moore and Cain (2007)				
Experiment 1	Quiz	Yes	University students	Yes
Experiment 2	Quiz	Yes	University students	Yes
Grossman and Owens (2012)				
	Quiz	Yes	University students and staffs	Noisy
Ryvkin, Krajc, and Ortmann (2012)				
Study 1	Score in exams	Yes	University students	Yes
Task 1 in Study 2	Math	Yes	University students	Yes
Task 2 in Study 2	Quiz	Yes	University students	Yes
Sabater-Grande, Nikolaos Georgantzís and Herranz-Zarzoso (2023)				
Experiment	Score in exams	Yes	University students	Yes

Table 2: Summary Statistics

	Control (N=31)		Exp (N=34)		Info (N=30)		InfoExp (N=33)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Male	0.42	0.50	0.47	0.51	0.47	0.51	0.48	0.51
Age	20.45	1.69	20.56	1.52	20.50	1.11	20.88	1.22
Job	0.84	0.37	0.76	0.43	0.93	0.25	0.85	0.36
Math	0.10	0.30	0.06	0.24	0.20	0.41	0.09	0.29
Loss propensity	2.77	1.45	3.12	1.65	3.43	1.59	3.50	1.93
Extraversion	3.84	1.49	3.62	1.20	3.85	1.40	4.36	1.64
Agreeableness	4.95	1.03	4.68	1.27	4.72	1.72	5.27	1.13
Conscientiousness	3.24	1.26	2.97	1.07	3.12	1.42	3.45	1.35
Neuroticism	4.82	1.14	4.66	1.35	4.22	1.45	4.34	1.48
Openness	3.95	1.25	4.03	1.39	3.90	1.43	4.00	1.25

Note: The p-values of the F-tests for the balance test are all greater than 0.1.

Table 3: Big Five

	I see myself as;	Strongly	Moderately	A Little	Nor Disagree	A Little	Moderately	Strongly
A	Extraverted, enthusiastic.	1	2	3	4	5	6	7
B	Critical, quarrelsome.	1	2	3	4	5	6	7
C	Dependable, self-disciplined.	1	2	3	4	5	6	7
D	Anxious, easily upset.	1	2	3	4	5	6	7
E	Open to new experiences, complex.	1	2	3	4	5	6	7
F	Reserved, quiet.	1	2	3	4	5	6	7
G	Sympathetic, warm.	1	2	3	4	5	6	7
H	Disorganized, careless.	1	2	3	4	5	6	7
I	Calm, emotionally stable.	1	2	3	4	5	6	7
J	Conventional, uncreative.	1	2	3	4	5	6	7

Table 4: Coin-toss Game

Number	Question	Chice A	Chice B
1	Heads: -100 YEN; Tails: +600 YEN	Participate	Not Participate
2	Heads: -200 YEN; Tails: +600 YEN	Participate	Not Participate
3	Heads: -300 YEN; Tails: +600 YEN	Participate	Not Participate
4	Heads: -400 YEN; Tails: +600 YEN	Participate	Not Participate
5	Heads: -500 YEN; Tails: +600 YEN	Participate	Not Participate
6	Heads: -600 YEN; Tails: +600 YEN	Participate	Not Participate
7	Heads: -700 YEN; Tails: +600 YEN	Participate	Not Participate

Table 5: Summary of the Results

	Control (N=31)		Exp (N=34)		Info (N=30)		InfoExp (N=33)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Answer	25.81	5.75	28.03	5.58	26.50	4.81	32.03	7.71
Goal	40.48	14.36	24.94	6.21	24.30	4.65	27.70	6.78
Success	0.23	0.43	0.76	0.43	0.70	0.47	0.88	0.33

Note: Answer represents the number of correct answers. Goal represents the goal set by subjects themselves. Success takes the value one if Answer is equal to or greater than Goal, otherwise zero.

Table 6: Pairwise Comparisons for Groups

				p-values			
Comparison	Variable	Difference	Effect Size	t-test	Ranksum	Variance	Exact
Exp - Control	Answer	2.223	0.393	0.119	0.124	0.868	-
Exp - Control	Goal	-15.543	-1.428	0.000	0.000	0.000	-
Exp - Control	Success	0.539	0.524	-	-	-	0.000
Info - Control	Answer	0.694	0.131	0.611	0.593	0.342	-
Info - Control	Goal	-16.184	-1.506	0.000	0.000	0.000	-
Info - Control	Success	0.474	0.458	-	-	-	0.000
InfoExp - Control	Answer	6.224	0.911	0.001	0.001	0.109	-
InfoExp - Control	Goal	-12.787	-1.150	0.000	0.000	0.000	-
InfoExp - Control	Success	0.653	0.646	-	-	-	0.000
Info - Exp	Answer	-1.529	-0.292	0.244	0.182	0.419	-
Info - Exp	Goal	-0.641	-0.116	0.639	0.968	0.117	-
Info - Exp	Success	-0.065	0.000	-	-	-	0.584
InfoExp - Exp	Answer	4.001	0.596	0.018	0.031	0.069	-
InfoExp - Exp	Goal	2.756	0.424	0.088	0.063	0.614	-
InfoExp - Exp	Success	0.114	0.084	-	-	-	0.340
InfoExp - Info	Answer	5.530	0.851	0.001	0.002	0.012	-
InfoExp - Info	Goal	3.397	0.579	0.023	0.018	0.043	-
InfoExp - Info	Success	0.179	0.180	-	-	-	0.120

Note: Effect size is cohen's d (for t-test) and phi (for fisher's exact test). 't-test' means Welch's t-test. 'Ranksum' means Wilcoxson ranksum test. 'Variance' means F-test of equality of variances. 'Exact' means Fisher's exact test.

Table 7: Regressions

	Answer		Goal		Success	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	25.806	19.895	40.484	56.094	0.226	−0.436
	(1.032)	(7.769)	(2.580)	(13.816)	(0.076)	(0.768)
Exp	2.223	2.269	−15.543	−15.730	0.539	0.552
	(1.408)	(1.489)	(2.791)	(2.752)	(0.106)	(0.110)
Info	0.694	0.361	−16.184	−17.814	0.474	0.508
	(1.355)	(1.501)	(2.716)	(2.727)	(0.114)	(0.121)
InfoExp	6.224	6.322	−12.787	−12.991	0.653	0.660
	(1.693)	(1.795)	(2.837)	(2.827)	(0.096)	(0.109)
Male		−1.983		−1.896		−0.057
		(1.216)		(1.708)		(0.088)
Age		0.360		−0.344		0.032
		(0.342)		(0.681)		(0.036)
Job		−0.304		−0.094		−0.029
		(1.204)		(2.341)		(0.124)
Math		1.159		4.831		0.034
		(2.028)		(2.754)		(0.131)
Loss propensity		0.124		0.239		−0.042
		(0.453)		(0.557)		(0.028)
Extraversion		0.008		0.563		0.009
		(0.443)		(0.518)		(0.024)
Agreeableness		−0.377		−1.190		0.009
		(0.410)		(0.515)		(0.031)
Conscientiousness		0.679		0.114		−0.012
		(0.454)		(0.742)		(0.034)
Neuroticism		−0.399		−1.374		0.007
		(0.430)		(0.525)		(0.033)
Openness		0.184		0.294		0.024
		(0.448)		(0.626)		(0.036)
Num.Obs.	128	126	128	126	128	126
R2	0.140	0.215	0.359	0.436	0.268	0.305

Note: Robust standard errors are in parentheses. Two samples are ommited due to missing values.

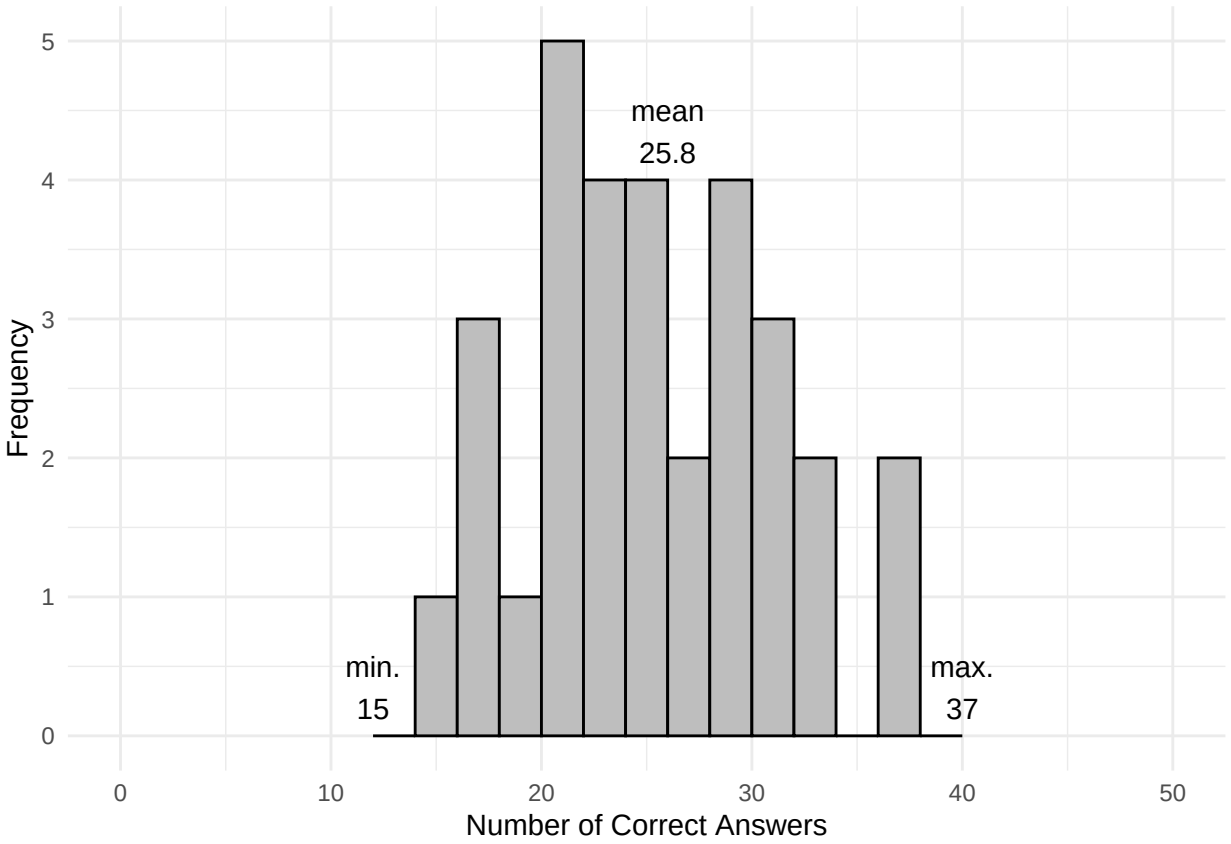

Figure 1: Number of correct answers in control group

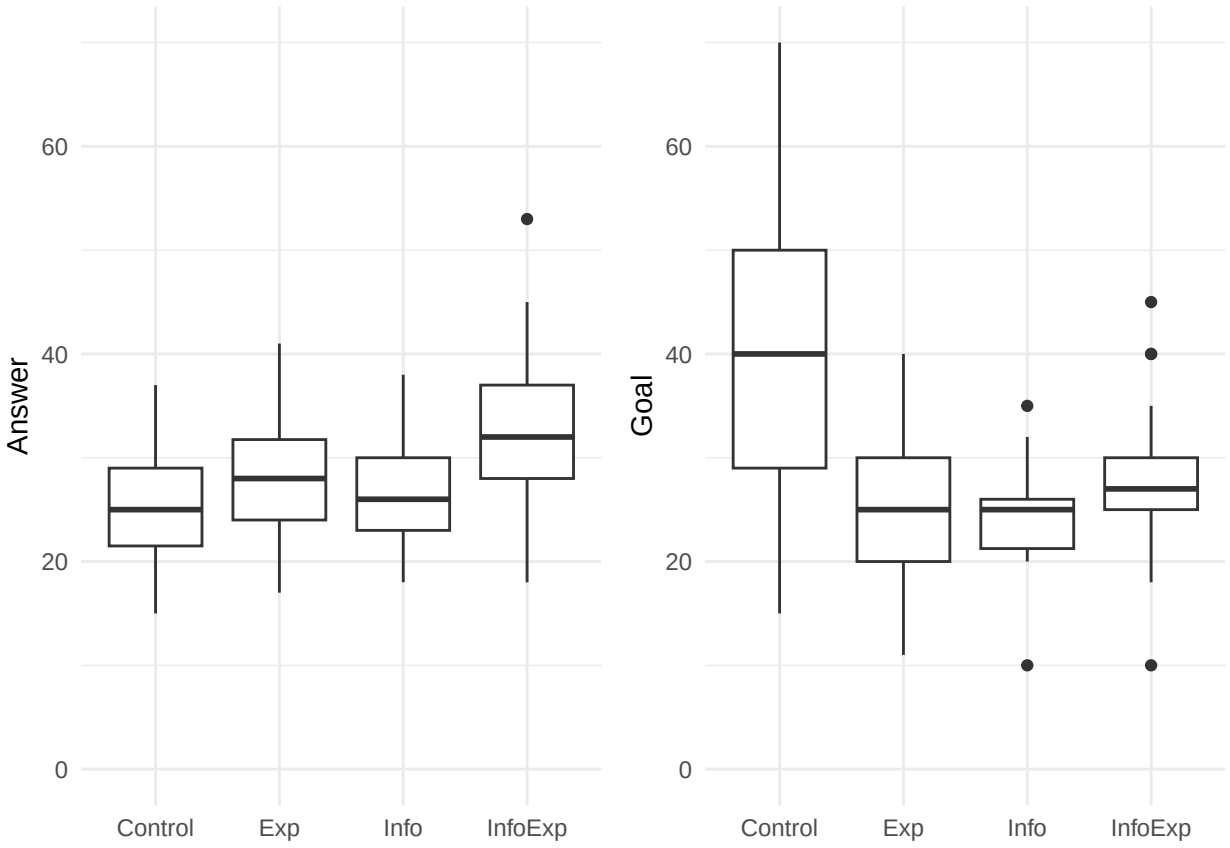

Figure 2: Boxplot of the number of Answer and Goal

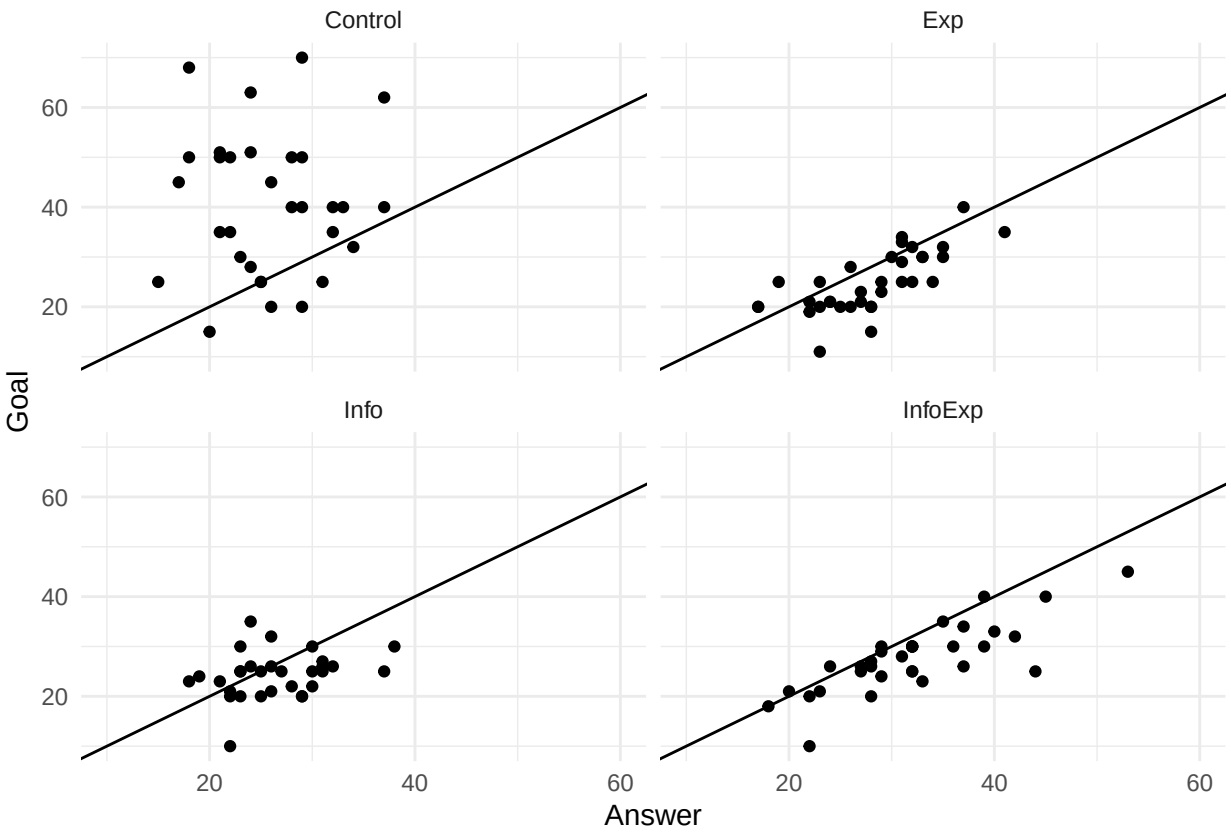

Figure 3: Scatterplots of Answer and Goal with 45 degree lines