Learning Outcomes, Feedback, and the Performance Effects of a Training Program

Christiane Hinerasky[a], René Fahr[a]

[a] University of Paderborn, Department Management, Warburger Str. 100, 33098 Paderborn, Germany

christiane.hinerasky@wiwi.upb.de, rene.fahr@wiwi.upb.de

Phone: +49 5251 603692 Fax: +49 5251 603550

Learning Outcomes, Feedback, and the Performance Effects of a Training Program

We empirically evaluate the job performance effects of a 6 month part-time training program in a large retail chain, in which participants received feedback on their training success based on a final exam. The data's quasi-experimental structure allows comparison of sales revenue of trained and untrained managers before, during and after the training. We find that the training significantly increases sales of very successful participants by approximately 8% during and by 7% after the training period. However, the training has a substantial *negative* impact on the post-training sales performance of unsuccessful participants. The results indicate that the indirect motivational effects of training programs may be much more important than the direct effects of skill acquisition.

Key Words: Evaluation, Returns to Company Training, Performance Effects, Feedback, Motivation

JEL Classification: C31, C33, J24, M53

Introduction

Even in the recent economic downturn, companies have spent substantial amounts on employee trainings. For instance, total company training expenditures in 2010 have been estimated at $52.8 billion in the US, an average of $1,041 per learner, with employees on average spending 40.1 hours per year in training (Training Magazine, 2010). Despite the considerable amounts of money that companies spend on training programs and despite the now large number of academic studies on the impact of training, firms rarely assess training outcomes on productivity and motivation comprehensively.

One reason for the lack of proper training evaluations is the often missing availability of individual objective performance data before and after training participation. While performance data are available on an organizational level, they typically cannot be attributed to individual training participants. Moreover, training may have very different effects on employees depending on their prior knowledge and individual predispositions, such as talents and preferences. Even within the same firm and occupation, employees may strongly differ in their ability and motivation to make use of the offered training.

In the language of Kirkpatrick's (1979) classical taxonomy of training evaluation criteria, firms therefore often rely on an assessment of either only the participant's *reaction*, i.e., their liking of the training program, or the *learning* defined as the "principles, facts, and techniques understood and absorbed" by the participants (Kirkpatrick and Kirkpatrick, 2006). While the former can, for instance, be easily assessed with a survey, the latter is typically measured by an exam at the end of the training. Assessing the impact on the participants' behavior and, moreover, on the (for instance financial) results for the company has proven to be much harder in practice.[1]

Individual examination results at the end of a training course typically do not reveal to what extent

[1] Sugrue and Rivera (2005), for instance, surveyed the American Society for Training and Development Benchmarking Forum and determine to what extent Kirkpatrick's four levels (reactions, learning, behavior, and results) are assessed in organizations. They found that while "reactions" were assessed in 91.3% of the cases and "learning" in 53.9%, the percentage dropped to 22.9% for "behavior" and only 7.6% for "results".

the training content has been transferred into the workplace. However, they can show whether a participant was able to apprehend the training content. It is often understood that there is a causal positive connection between the levels of training evaluation. For instance, a higher degree of "learning" should indicate more human capital formation, in turn leading to different and more productive "behavior" and, in turn, better business "results".[2] However, as we will argue in the following, the link between the individual learning outcomes and the returns to training might be more intricate. When employers try to measure individual learning outcomes, the measurement also provides direct feedback to the employee. This feedback may also affect the way in which she or he reacts to training participation beyond the direct learning effects. In particular, demoralization resulting from a negative performance feedback in the exit exam of a training course might affect performance at the workplace negatively. In this case, potential de-motivational effects may even counteract any positive effects derived from human capital formation.

As many employer-provided training programs typically last for long periods of time while participants continue working in their jobs, performance may be affected through very different channels. Beyond the increase in skills and knowledge directly transferred, trainings may improve the identification with the employer and in turn employee motivation. While the first effect should lead to a higher productivity in future periods, the second effect may already have an impact in the short term. Hence, it seems important to study the *timing* of potential productivity effects during as well as after training.

In this paper we therefore make use of a unique data set to investigate (i) how productivity effects of a training program change over time during and after training participation, (ii) how public feedback influences on-the-job productivity and (iii) how productivity effects are measured taking the heterogeneity of the learner, according to high or low performance in the exam by means of individual learning outcomes, into account.

We use data from a retail company with more than 500 stores, which organized a 6 month part-time

[2] See, for instance, Alliger and Janak (1989) for a critical discussion of this assumption. Alliger et al. (1997) conducted a meta-analysis investigating the relations between the different levels.

training program for the store managers. A precise performance measure (sales revenue) is observed for each store in each month over the four consecutive years before, during, and after training, allowing the observation of transferability and the behavior of performance effects over a comparatively long period of time. In addition, the panel data offer detailed store and individual specific information that help to control for heterogeneity on top of common panel data estimation techniques. Another key feature of the data set is that the learning outcomes are measured directly after the end of the program because participants had to take part in a graded exam. The public distribution of certificates and grades enables us to be among the first to analyze how learning feedback can influence on-the-job performance.

There are now a substantial number of studies on the effects of training programs in different fields, from psychology to economics. In recent years, meta-analyses have aggregated the results of numerous individual studies. Arthur et al. (2003), for instance, conducts a comprehensive meta-analysis on the effectiveness of training programs, finding medium to large effect sizes. Bartel (1989) analyzed the effect of company training on firm productivity and finds no effect once individual and company specific information was controlled for. Similarly, Black and Lynch (1996) do not find a significant impact of the number of trained employees on a company's productivity, and Barrett and O'Connell (2001) find no effect for "specific" training. Several studies on the other hand show a positive and significant impact. These are, among others, Bartel (1995), Bishop (1990), Holzer et al. (1993), Barron, Berger, and Black (1998), Delaney and Huselid (1996), Dearden, Reed, and VanReenen (2006) and Liu and Batt (2007). Morrow, Jarrett, and Rupinski (1997) use a case study framework similar to ours to evaluate a pharmaceutical company's training programs and find technical and sales training to have higher impact than managerial training. However, these studies have not analyzed the extent to which a measure of the learning which is communicated to the training participants affects the results of the training program. From a rather qualitative perspective, Goodman, Wood, and Chen (2011) assess how trainees use and interpret feedback information received during training and how the specificity of feedback influences the amounts of learning and transfer. However, to our knowledge, only one study provides a comprehensive theoretical perspective on how

learning outcomes may be assessed and used for training evaluation (Kraiger, Ford, and Salas, 1993).

Most of these studies, furthermore, can only observe a maximum of a few months after the training has been finished and therefore cannot assess how the effects evolve over time. As Mathieu, Tannenbaum, and Salas (1992) have remarked, the post-training period, especially the trainee's reaction to training, plays an important moderating role between training motivation and learning outcomes; i.e., positive reactions to training can lead to higher posttest performance scores. Despite the works of Noe (1986) and Noe and Schmitt (1986) who emphasize the motivational and environmental influences on training effectiveness, the transferability of training in empirical studies is only imperfectly observed.

The paper proceeds as follows. First the data, company background and training program are presented, including theoretical considerations and their operationalization. The next section gives key results and their discussion. The last section concludes.

The Data

The data for this study was collected from the company's personnel records and financial data for the time between January 2004 and December 2007.[3] From the company's 500 subsidiaries, 340 were managed by one supervisor during the entire 4-year period and therefore have been selected for observation. Among these, 165 store managers took part in the training program during the period of observation, while 175 store managers did not. The average training participant (non-participant) is 41 (44) years old, has 3.3 (3.2) full-time employees, has approximately 10.7 (10.9) years of experience with the company, and is leading a store that existed for 11.4 (11.5) years. Although participation is not being determined purely at random, both groups are surprisingly similar. Each store can be sorted into one of three different sales level categories, which depend on its size, location, and regional purchasing power. As an objective productivity

[3] The data have been used before in Hinerasky and Fahr (2014).

measure, monthly sales revenues are observed for each store during the entire period of 48 months. This results in a total sample of 16,320 observations. For each store, additional information is available on the number of employees and their respective hours of work, the manager's age and tenure, the age of the store, a dummy indicating whether the store is a franchise store, as well as numerous time-invariant characteristics further described in Table 1.

Table 1: Description of Variables and Descriptive Statistics

Individual time-variant Variables	**Obs.**	**Mean**	**Std. Dev.**
Age	13056	43.18	8.41
Tenure	13008	11.06	6.45
Existence	16320	13.26	5.72
	Obs.	**Percent**	**Cum.**
Training Participation:	16320		
Participation *(everT=1)*	7920		
While Training *(inT=1)*	869		
After Training *(afterT=1)*	1364		
Franchise Branch:	15936		
Yes *(partner=1)*	3,304	20.73	20.73
No	12,632	79.27	100

Store – specific Variables	**Obs.**	**Mean**	**Std. Dev.**
Sales Revenue *(sales)*	16320	145.84	40.49
Full Time Equivalents *(fullemp)*	10356	3.26	0.94
	Obs.	**Percent**	**Cum.**
Store Environment:	16320		
Shopping Center *(standcenter)*	6,384	39.12	39.12
Downtown *(standinner)*	2,976	18.24	57.35
Stand Alone *(standalone)*	6,960	42.65	100.00
Store Environment Detail:	16320		
1a-Location *(stand1a)*	1,200	7.35	7.35
1b-Location *(stand1b)*	1,776	10.88	18.24
< 10 tsqm floor space & discount *(standun10tqmdisc)*	2,544	15.59	33.82
< 10 tsqm floor space & market *(standun10tqmretail)*	768	4.71	38.53
> 10 tsqm floor space & discount *(standov10tqmdisc)*	1,536	9.41	47.94
> 10 tsqm floor space & market *(standov10tqmretail)*	1,536	9.41	57.35
Weak commerce *(standtypp7)*	3,840	23.53	80.88

Strong commerce *(standtypp8)*	3,120	19.12	100.00
Sales Revenue Category:	13056		
Small *(salesun1m)*	9,216	70.59	70.59
Medium *(salesbet12m)*	2,592	19.85	80.15
Large *(salesov2m)*	1,248	19.85	100
Region:	16320		
West Germany *(westger=1)*	10032	61.47	61.47
East Germany	6288	38.53	100
Federal State:	16320		
Baden-Württemberg *(bw)*	3,024	18.53	18.53
Bayern *(ba)*	2,496	15.29	33.82
Berlin *(ber)*	528	3.24	37.06
Brandenburg *(bra)*	1,008	6.18	43.24
Bremen *(bre)*	144	0.88	44.12
Hamburg *(ha)*	96	0.59	44.71
Hessen *(he)*	576	3.53	48.24
Mecklenburg-Vorpommern *(mv)*	720	4.41	52.65
Niedersachsen *(ns)*	720	4.41	57.06
Nordrhein-Westfalen *(nw)*	1,824	11.18	68.24
Rheinland-Pfalz *(rp)*	768	4.71	72.94
Saarland *(sa)*	240	1.47	74.41
Sachsen *(sac)*	1,872	11.47	85.88
Sachsen-Anhalt *(saa)*	1,152	7.06	92.94
Schleswig-Holstein *(sh)*	144	0.88	93.82
Thüringen *(th)*	1,008	6.18	100.00
Size of Town:	16320		
Provincial Town < 20,000 inhabitants *(townsmall)*	8,544	52.35	52.35
Middle Town 20,000 - 100,000 inhabitants *(townmid)*	4,512	27.65	80.00
Major City > 100,000 inhabitants *(townbig)*	3,264	20.00	100.00
Branch Management:	13056		
Line Manager *(ffk=1)*	12288	94.12	94.12
Deputy Manager *(supffk=1)*	48	0.37	94.49
Supervising Employee *(empffk=1)*	720	5.51	100
Regional Executive Manager:	16320		
Dummy for each Executive Manager *(mm1-mm3)*			
District:	16320		
Dummy for each of 23 districts *(regioncode1-23)*			

Note: Certain information (e.g. age, tenure, number of full-time employees) is not available for franchise stores, which amount to 3,304 of 16,320 observations.

Company training, human capital and feedback

Background

The retail company under investigation sells products to private customers in the lower and middle price segment in an industry with tough price competition. Although a wide range of different products are offered, the product portfolio is very similar across all 500 stores. The training program was introduced in 2005 as part of a change in marketing strategy, intended to shift the company away from its former focus on low prices to a stronger orientation towards quality and service. The major aim of the program was to teach store managers business and marketing-related knowledge, which they in turn should communicate to their employees and thereby improve the company image and sales. Therefore, only store managers were eligible for participation. Training is built as an e-Learning tool with audio and video tutorials and covers subjects such as time management, purchasing, sales, organization, leadership, marketing, law, and managerial-economics. During a 6 month period, participants spent an average of 10 hours per week on studying and preparing courses supplemental to their regular work load. Four times during the training phase, participants virtually met in chat groups with a senior professional to compare notes and to foster an active discussion. In addition, trainees were able to communicate via chat at all times. The program was initiated in October 2005, and within the observation period store managers in 10 cohorts went through training at different points in time starting in October 2005, March, June, September and November 2006, and February, March, April, September and October 2007.

Initially training participation was voluntary. Admission to training was based on a first-come, first-served strategy. However, all store managers are asked to participate in the training at some point in time. Courses are passed upon the accomplishment of a final exam that consists of a written and oral test held by the Chamber of Commerce and Industry (CCI). Afterwards participants are rewarded with a CCI-certificate. The written test covers all learning material, asking 120 multiple choice questions. In the oral test, participants are faced with different day-to-day issues, e.g., complaining customers or a low sales volume,

and have to deliver problem solving suggestions and their recommended course of action in front of an examination board. Ninety-nine percent of the participants passed the final exam obtaining grade 4(D) or higher (grades range from 1 (very good) to 5 (unsatisfactory)), though a few needed a second try. Finishing the course is rewarded with a monthly pay raise of 100 Euro, and those who passed with distinction (grades 1(A) and 2(B)), on top, were refunded the entire training costs of 260 Euro. Within the first three years (2005 to 2007), 165 branch managers took part in groups of 20 to 32 participants (∅ 26.1) and obtained the following average grades, as observed in Table 2.

At an official ceremony, certificates are handed over to each course participant, especially honoring very good managers. As each person is called by name and his respective grade, participants not only receive feedback on their own performance but also on the achievement of all of the other participants.[4] This allows estimation of training effects with respect to the grade and hence the feedback received.

Table 2: Feedback Measure, Grading of final exam for all courses in 2004-2007

	Written Exam		Oral Exam		Final Grade	
Grade	Freq./Groups	%	Freq./Groups	%	Freq./Groups	%
1 = A	768/16	10.00	2,256/47	29.38	720/15	9.38
2 = B	3,600/75	46.88	1,776/37	23.12	3,408/71	44.38
3 = C	3,120/65	40.62	1,728/36	22.50	2,880/60	37.50
4 = D	192/4	2.50	1,344/28	17.50	624/13	8.12
5 = E	-	-	528/11	6.88	-	-
6 = F	-	-	48/1	0.62	48/1	0.62
Total	7680/160	100.00	7680/160	100.00	7680/160	100.00

Note: Grades are given on a standardized basis: 100-92 Points Grade 1(A); <92-81 Points Grade 2(B); <81-67 Points Grade 3(C); <67-50 Points Grade 4(C); <50-30 Points Grade 5(E); <30-0 Points Grade 6(F).

Theoretical Considerations

During-Training Performance

Skills should be acquired continuously over the training period and, in turn, according to human capital

[4] This information has been obtained from the participation of one of the authors at the final examination and the subsequent official ceremony.

theory (Becker, 1964), productivity should already rise as well during the program. A key benefit of the e-Learning training form should be that the training content of the prior evening may be directly used or practiced the next day during working hours, as opportunities for practice are a crucial point in the transfer of training (Olivero, Bane, and Kopelman, 1997). Additionally, the discussion in chat rooms and sporadic coaching can facilitate the transfer of training, especially if the "coaching fosters the development and use of knowledge imparted during training" (Olivero, Bane, and Kopelman, 1997, p. 461)[5]. However, training which is held in addition to regular work schedules might overburden less able branch managers. Working with the e-Learning training program in evening hours might lead those participants who are already very burdened in their jobs to exert less effort during work hours and hence to underperform compared to their pre-training performance. More able managers on the other hand might be motivated by the possibility to reflect on day-to-day issues using the training tool. Furthermore, they might be able to directly transfer training content into their daily routine and thereby increase performance during training participation compared to pre-participation performance. We expect the productivity among able managers to increase during training because of the relatively long training period, immediate transfer possibilities, and sporadic coaching by a firm professional.

Feedback and After-Training Performance

A key feature of the data set is that the training ends with an exam providing training participants with feedback on their training performance. Hence, performance should be affected after the end of the training because (i) increased human capital should lead to increased productivity given the transferability of knowledge. However, furthermore, (ii) the feedback information contained in the exam results may affect task performance after training. In the same way that the identification of employees as "talents" can have positive effects on their commitment to increase performance (Björkman et al., 2013), this feedback effect

[5] More recent work on the effects of coaching and its underlying theoretical model can be found in Gray (2006), Agarwal, Angst and Magni (2009), or Liu and Batt (2010).

can be positive as it increases individual motivation or helps employees to identify strengths and weaknesses. This is especially the case when feedback causes internal, social, or external consequences (Kopelman, 1982) and when it allows individuals to establish specific goals (Locke et al., 1981). Feedback is also demonstrated to improve task performance when it is combined with reflection (Anseel, Lievens, and Schollaert, 2009). However, as has been stressed in this literature, when feedback is followed by no or only little (psychological) consequences, the effects of it can be minimal. This is also the case when the task to be fulfilled is perceived as not meaningful and without responsibility (Kopelman, 1986). Montague and Webber (1965) find in a visual monitoring laboratory experiment that nonspecific feedback, e.g., categorization into superior, adequate, or poor performance, only has small additional effects on task performance. Kluger and DeNisi (1996) also find high variability of feedback interventions on performance in a large scale meta-analysis. While a feedback intervention typically improves performance, in one-third of the cases it reduced performance. Reviewing 72 field and laboratory studies dealing with cognitive or motor work activities[6], Kopelman (1986) concludes "the person will experience feelings of increased self-approval and heightened self-regard upon performing well, and the opposite feelings upon performing poorly". A source for the observed variability of feedback effects is shown in McFarland and Miller (1994) and indicates that persons respond to feedback with different performance levels depending on the emphasis of positive or negative feedback framing. Negative feedback impedes performance among pessimistic types while it increases performance among optimistic individuals. Sitzmann and Johnson (2012) also show that a lack of alignment between trainees' self-assessment of knowledge and their actual performance impacts subsequent performance and even skill attrition. Negative reactions after training participation are likely to occur when the trainees' self-assessment overestimates their true performance.

Moreover, it has been stressed that public feedback, as used in the firm under study, where participants are aware of their individual and others' results, alters performance by causing competition

[6] Among the 72 studies, none provide learning feedback.

among the participants. Runnion, Watson, and McWhorter (1978) and Nordstrom, Lorenzi, and Hall (1990), for instance, claim that as participants opt to avoid the possible embarrassment that may accompany poor performance, their performance can be additionally improved by publicly available results. More recently, small-scale studies have taken a closer look at public feedback in housekeeping (Leivo, 2001), sanitizing practices among Certified Nurse Anesthetists (Stephens and Ludwig, 2005) and safe driving among pizza deliverers (Ludwig et al., 2001) and documented its impact. To our knowledge, however, there has not been a quantitative study on the effects of public feedback on performance in a learning context. Public feedback incentivizes good and very good grades as it aims at and rewards high performance in the final exam. Conveying these results to a training context, we assume that participants respond to positive feedback with increased performance and to negative feedback with decreased performance.

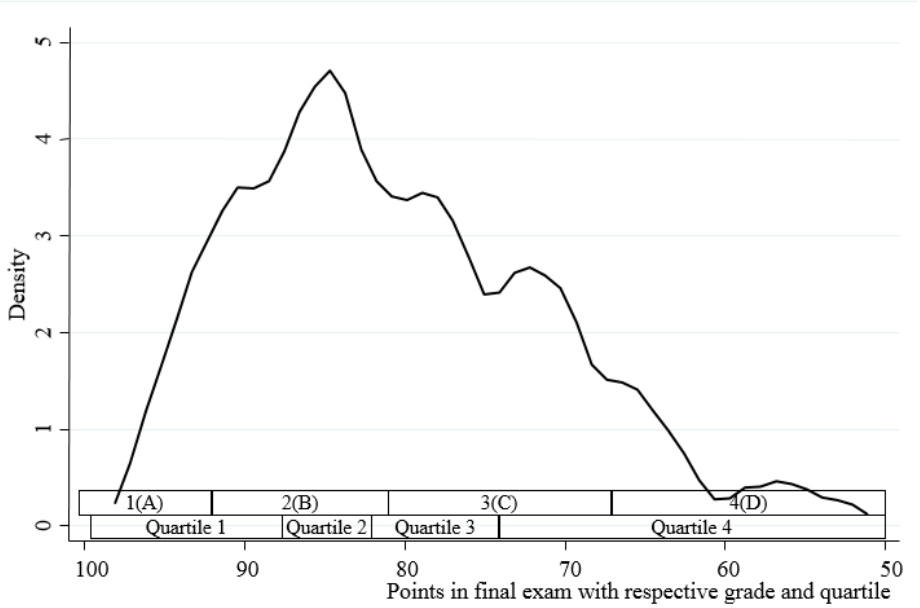
Identification Strategy

It is important to note that potential positive productivity effects *after* the training may stem from either increased skills and knowledge or the motivational effects of the feedback provided. The structure of our data allows us to disentangle these effects by using two different approaches. First, analysis of the during-training period identifies whether knowledge and skills have been built and transferred before participants even took part in the exam. Second, strong discouraging effects from negative feedback can be identified when post-training performance of those receiving negative feedback falls below the pre- and within-training outcomes.

Second, the particular structure in which the learning outcomes are measured and communicated to the training participants allows for another empirical approach to identify positive or potentially negative feedback effects. In particular, we calculate quartiles based on the distribution of total points a participant has reached in both his written and oral examination. The points achieved proxy the amount of skills and knowledge built up as participants who reach full score in the final exam display a higher stock of skills

than those with, e.g., only 20 % (Human Capital Effect). A participant who finishes with a final score that places him in quartile 1 is assumed to have accumulated a higher stock of human capital than participants in quartiles 2, 3 or 4. Figure 1 breaks down the quartiles with the respective grades and shows that, e.g., a participant with grade 2(B) can appear in quartile 1 as well as 2 or 3. This is because 44% of all participants receive a 2(B) or better. A participant who finishes with 3(C) reaches a score that places him either in quartile 3 or 4.[7] This approach makes it possible to disentangle human capital effects from feedbacks effects whereby the stock of human capital is proxied by the quartile reached and feedback by given grade. In particular, finding negative training effects for training variables interacted with lower grades, but not for training variables interacted with lower quartiles, would suggest the presence of negative feedback effects.

Figure 1: Distribution of Grades and Quartiles



Note: Grades are given on a standardized basis: 100-92 Points Grade 1(A); <92-81 Points Grade 2(B); <81-67 Points Grade 3(C); <67-50 Points Grade 4(C); <50-30 Points Grade 5(E); <30-0 Points Grade 6(F). Quartiles are calculated on the participant's overall score/ percentile in the final exam. Quartile 1 (n=1,968) contains 41 groups, quartile 2 (n=2,016) contains 42 groups, quartile 3 (n=1,920) contains 40 groups, quartile 4 (n=2,016) contains 42 groups.

[7] A cross-tabulation of final grades and quartiles as in Table A1 shows the numeric overlap of both categories.

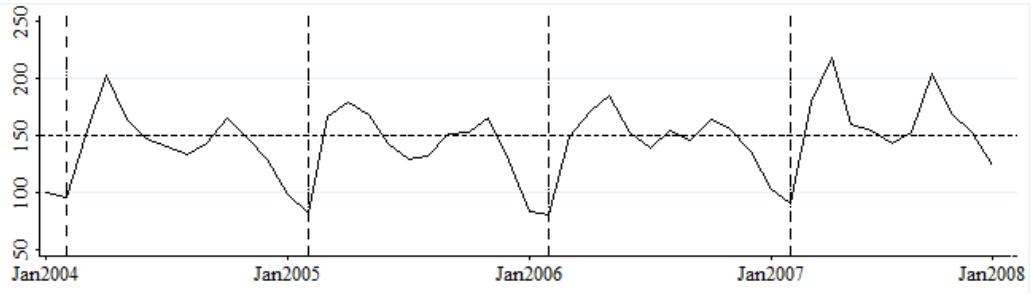
Operationalization

We measure training participation with two dummy variables: *inT* and *afterT*. *inT* takes the value 1 if the respective store manager is taking part in training in the month under consideration and *afterT* takes the value 1 for all months after the respective store manager has finished the program. A store manager's productivity is measured by his respective store's sales revenue $y_{i,t}$ in month t for store i. Due to privacy restrictions, monthly sales are normalized by dividing a store's current sales revenue by its sales revenue in January 2004. The performance measure y for store i in January 2004 is set to 100 and, hence, equals $y_{i,Jan2004} = 100$. Performance measure y for store i in month t is therefore

$$y_{i,t} = \left(Sales_{i,t} / Sales_{i,Jan2004} \right) * 100 \quad (1)$$

As $y_{i,t}$ corresponds to the sales development for an individual and store level beginning with January 2004, a comparison between different stores controlling for different prior sales levels is no longer possible. However, this feature can be neglected as our research question is concerned with an investigation of the individual reaction to training, i.e., each individual performance trend during and after training compared to before. It is therefore the within- rather than the between- variation that is required for our analysis.

Figure 2: Normalized Company Sales Revenue for 2004 – 2008



Note: Normalized non-deflated sales revenue is calculated as the average of all store sales revenues. Each year's all-time low in February is labeled by vertical lines.

Figure 2 provides an overview of the company's overall sales revenue pattern calculated as the average of

the observed stores' revenues. Spring and autumn are peak seasons with average sales revenues of approximately 170% of the January 2004 sales levels. Sales during summer and winter regularly decrease following the same pattern in each year. While the overall sales level is stable at approximately 150% for the years 2004, 2005, and 2006, it has increased substantially in 2007. To account for these patterns and to eliminate any seasonal effects, we model the seasonal structure by including dummies for years and months as suggested by Hanssens, Parsons, and Schultz (2003). Due to the low variation in sales during the years 2005 and 2006, we include a dummy for those two years and one for the year 2007 that accounts for higher revenue in the retail market (see Figure 2). Because normalizing each store's sales revenue set the dependent variable to 100 in the base month, any variation between stores was eliminated. We therefore estimate the within-variation with a fixed effects model that accounts for time invariant idiosyncratic effects in the observed sales of training participants. We test for autocorrelation of the error term, which is a common concern in time series, using Wooldridge's (2002) test for panel data. Drukker (2003) using Monte Carlo simulations showed good size and power properties of Wooldridge's (2002) test in samples of reasonable size for all common estimation procedures. As serial correlation of the error term can cause inefficient estimates, we calculate models with first-order autoregressive disturbance terms. Recent empirical literature suggest estimating models with clustered standard errors for data with a possible group structure, such as, e.g., correlated outcome variables (Angrist and Pischke, 2009). Nevertheless, as our outcome variable *sales revenue* is not simply correlated over time in some way but clearly follows a path dependency, a more conservative approach would be to estimate an autoregressive model that determines the form of how idiosyncratic errors are correlated over time, rather than by just assuming some sort of correlation as would be done by a model with clustered standard errors. Hence, we estimate cross-sectional time-series fixed effects models with a first-order autoregressive disturbance term and account for time invariant idiosyncratic effects in the observed sales of training participants.

To assess how productivity effects differ among employees, we estimate interaction effects for two sub-groups: the high and low performers. High performers received a grade of 2(B) or higher in the final

exam whereas low performers finished with 3(C) or worse. Among all of the participants, 56.88% passed the written exam with grades 1(A) or 2(B). In the oral exam, this number is reduced to 52.50%, where almost one-third (29.4%) passed with grade 1 and 23.1% with grade 2. Overall, 9.4% completed the training with grade 1 and 44.4% with grade 2. Therefore, approximately 53.75% of all participants received a reimbursement of their training costs. Average final grades and points separated by cohort can be found in Table A2 in the appendix.

Empirical Results

Because the training program lasts 6 months and participation allows those enrolled in the program to follow their regular job, we control with *inT* for time periods during training and with *afterT* for the time periods after training participation in the estimates. A list of variables used to control for branch-specifics and individual time-variants is displayed in Table 1. Regression results for the baseline and interacted regressions are reported in Table 3. For the average training participant, training does not lead to a significant increase in sales during the training phase (Specification 1 in Table 3). This can be explained by the high time effort participants have to invest for studying and the fact that training commonly shows its effect with a timely delay. However, surprisingly, we find a weak significant negative effect on time periods after training participation where sales revenues even undercut pre-training levels by 3.17 percentage points. However, this effect is strongly moderated by the results of the final exam, as observed from specifications (2) and (3). Specification (2) displays in-training and after-training effects with respect to different quartiles regarding the total points a participant attained in both his written and oral examination. Although this division can only be made after the exam has been taken, Table 3 shows a significant heterogeneity between participants already within the training phase.

Measuring performance outcomes and accounting for the heterogeneity of the learners, we find that participants in the first quartile increase their sales revenue in training significantly by 5.52 percentage points. This increase, however, is not persistent after the training. Even participants who finish in quartile 2

show no qualification effect during the training (the sum of the coefficient and the interaction term is close to zero) and a weakly significant negative effect after the training. Individual gain in quartile 3 does not differ strongly from that in quartile 1. However, most interestingly, participants in quartile 4 perform significantly worse during and after training. Apparently, as working days were filled with both work and study, sales revenues suffered during the training period for those participants who seemingly had difficulty in learning the contents of the training program. Moreover, this effect seems to be persistent after the end of the course, which is explored in more detail below. Therefore, the training does not have a general negative long-term effect on sales revenues. Negative effects are rather driven by participants who either did not get along with the training structure or its content. Participants, however, who actively used the training as indicated by their training results are able to transfer the training contents to their jobs.

Table 3: Regression Results with Interactions: Quartiles (Qualification Effect) & Grades (Feedback Effect)

	(1)	(2)		(3)
SALES	FE AR(1)	FE AR(1)		FE AR(1)
	Baseline	**Quartiles**		**Grades**
inT	0.7254	5.5206**	inT	8.1245**
	(1.44)	(2.50)		(4.01)
afterT	-3.1653*	1.2142	afterT	7.2861**
	(1.72)	(2.50)		(3.52)
inTquartile2		-6.0422*	inTgrade2	-7.1734*
		(3.11)		(4.28)
inTquartile3		-3.6907	inTgrade3	-7.7899*
		(3.64)		(4.47)
inTquartile4		-8.7022***	inTgrade4	-13.8936**
		(3.34)		(5.41)
afterTquartile2		-5.2455*	afterTgrade2	-11.2814***
		(2.76)		(3.60)
afterTquartile3		-3.4434	afterTgrade3	-11.3870***
		(3.41)		(3.85)
afterTquartile4		-8.5827***	afterTgrade4	-16.0324***
		(3.14)		(5.07)
Dummys for month	Yes	Yes		Yes
Dummys for year	Yes	Yes		Yes
Individual time-variant variables	Yes	Yes		Yes
Store-specific variables	Yes	Yes		Yes
Observations/ groups	5004/108	5004/108		5004/108
Adj. R-squared	0.63	0.63		0.63

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Specifications (1)-(3) assume that the disturbance term is first-order autoregressive. One person with a final grade of 6(F) was excluded from all specifications. Reference group in

Specification (2) are participants in quartile 1 and in Specification (3) participants with grade 1(A). Results of Specification (3) are robust to a change of reference groups to the numeric largest group of grade 2(B). See Figure 1 to learn how quartiles relate to grades.

When studying possible effects that feedback on learning outcomes can have on participants, it is insightful to compare sales revenues before and after training participation. In Specification (3), post-training effects for each grade are calculated. Participants with grade 1(A) display positive post-training effects, which can be explained by either positive feedback that motivates higher effort or satisfaction from the fact that studying has paid off and that training fees are refunded. Grades 2, however, should be equally incentivized by the refund, but show a negative post-training effect of approximately 4 percentage points. This suggests that satisfaction from feedback about a very good grade exceeds the power of monetary incentives. This is even more evident for attendants who finish with 4(D) because they are not only unable to make use of the training content during training but even show strong de-motivational effects from negative feedback when sales revenue falls strongly below pre-training levels.

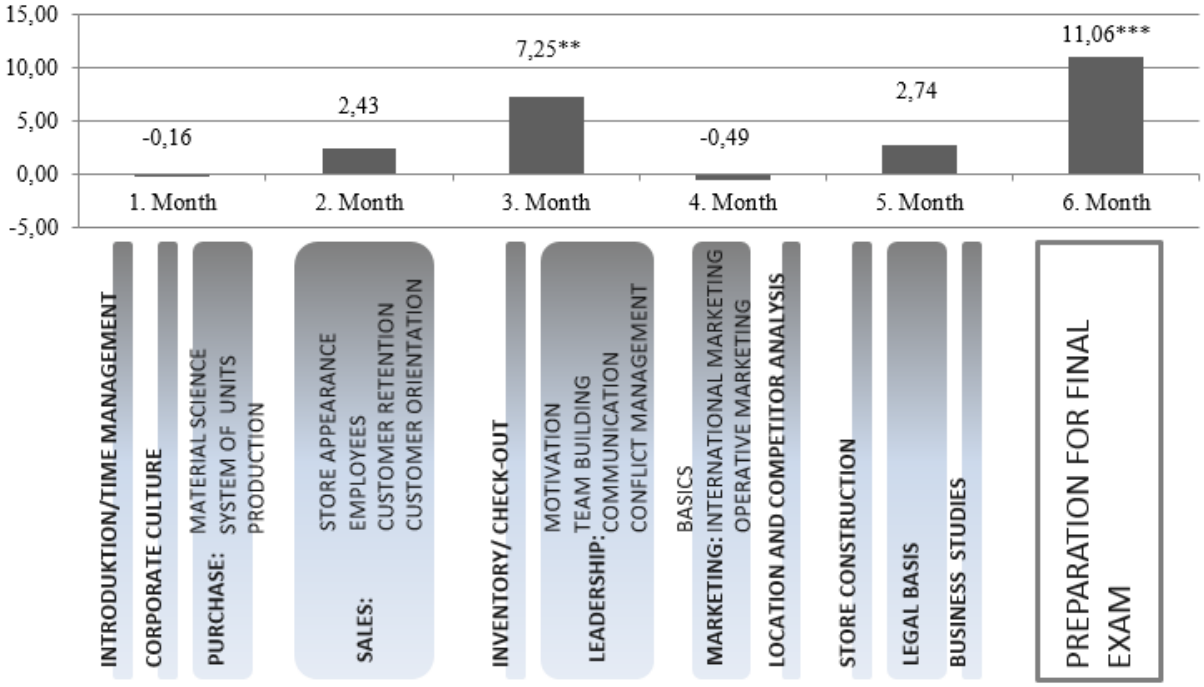
Table 4: Regression Results: Months during training in detail

SALES	(1) FE AR(1) **Baseline**	(2) FE AR(1) **High Performers**	(3) FE AR(1) **Low Performers**
1st month	-0.8370	-0.1509	-0.8793
	(2.08)	(2.69)	(3.30)
2nd month	-0.4702	2.4316	-4.2285
	(2.24)	(2.85)	(3.58)
3rd month	1.7881	7.2462**	-4.9347
	(2.29)	(2.92)	(3.67)
4th month	-2.8352	-0.4927	-5.5279
	(2.42)	(3.00)	(4.02)
5th month	0.8437	2.7447	-1.8420
	(2.53)	(3.14)	(4.22)
6th month	8.7134***	11.0589***	5.3485
	(2.74)	(3.35)	(4.71)
afterT	-2.3221	1.0978	-6.0477**
	(1.75)	(2.21)	(2.91)
Dummys for month	Yes	Yes	Yes
Dummys for year	Yes	Yes	Yes
Individual time-variant variables	Yes	Yes	Yes
Store-specific variables	Yes	Yes	Yes
Observations/groups	5004/108	2878/62	2126/46
Adj. R-squared	0.63	0.65	0.60

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Specifications (1)-(3) assume that the disturbance term is first-order autoregressive. One person with a final grade of 6(F) was excluded from all specifications.

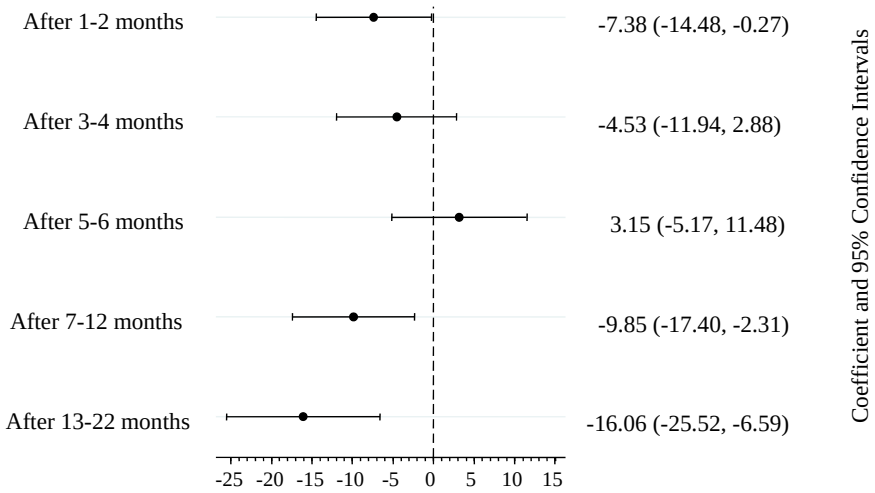
Because the training effect seems to vary highly over the course of training, we can make use of the detailed monthly information on sales and training content to further examine what happened during the entire training phase. Table 4 shows the results for the entire sample (Specification 1), for high performers who will receive grades 1(A) or 2(B) (Specification 2), and low performers who will end up with grade 3(C) or worse (Specification 3). Furthermore, Figure 3 displays both monthly training effects (for high performers) and the respective content of training graphically. High performers seem to benefit the most from training units in the third and sixth month, as apparent from significant revenue increases of 7.25 percentage points and 11.06 percentage points, respectively. Low performers, on the other hand, appear to not be able to immediately transfer any of the training content into the workplace. The topics covered in weeks 9-12 are inventory, general organization and cash management, and notably, topics on motivation, teambuilding, leadership and communication.

Figure 3: Monthly Training Effects and Training Content (only High Performers)



Apparently, these motivational issues bring about a rather strong short-term push in performance. During the sixth month, no training is provided and participants are free to study for the final exam in the evenings. As can be expected, effort, measured by daily time investments, typically increases once the exam approaches (Richard and Diefendorff, 2011). This phase of rehearsal apparently helps to consolidate the theoretical material provided by separate learning modules due to a strengthened focus on the entire course content. This fact also points to the importance of finishing courses with final exams because they lead participants to deepen and improve their knowledge and skills (see Figure 3). However, despite the positive effect that exam preparation has on some participants, eventual low performers will apparently suffer from the negative feedback given by their performance in the exam. Moreover, the negative feedback effect seems to persist after the end of the training. Figure 4 shows the time pattern for low performers in the months after participation.

Figure 4: Coefficients and Confidence Intervals from an AR(1) Fixed Effects Regression of months after participation on sales revenue for low-performers with grades 3(C) or 4(D)



Note: If the value 0 lies within the confidence interval of each independent variable as denoted by a dashed line, it is very likely that the independent variable has no impact on the dependent variable in the population. The number of observations in the categories after 1-2 months is 250, 3-4 is 238, 5-6 is 194, a7-12 is 384 and a13-22 is 288. Additional t-values for the variables after 1-2, after 7-12 and after 13-22 are 2.03, 2.56 and 3.32. Results are based on an AR(1) fixed effects regression with additional controls for individual time-variants and store-specifics (Table 1).

Sales revenue is significantly lower directly after participation in months 1 and 2, and again later in months 7 to 22. We conclude that negative feedback causes long-term discouragement, which can consistently lower sales revenues even months and years after participation. High performers on the contrary do not exhibit any negative effects in the post-training period.

Because we have reasonable amount of information on the store's organizational features, we also checked whether the training's short term effects unfold their impact faster in small stores, as insights in team leadership and team motivations may be implemented more quickly. However, there is no indication for such an effect in the data.

Discussion

The key result of the present study is that indirect motivational effects of training programs may be much more important than direct benefits through acquired skills and knowledge. Only high performers with very good grades were able to persistently increase their store's sales revenues due to the training participation. For low performers, the participation in a training program apparently has two types of detrimental effects. On the one hand, their sales performance goes down already during the training – potentially because they have to allocate time and effort from the regular sales task. For the high performers, this effect seems to be counterbalanced by immediate gains from learning. On the other hand, the low performers are in addition strongly de-motivated from the negative feedback on their training results.

De-motivational effects on low performers may be related to several reasons. Responding to training with no increase in productivity can simply be caused by a missing development of human capital. However, low performers respond with even lower sales revenue after training participation compared to before. This could be interpreted as human capital depreciation during the initial training phase, but this human capital depreciation would only be plausible had participants been released from their regular job during the 6-month training. In that case, participants had not only been taking part in an unsuccessful training program

but had also lost their job routine, which culminates in lower post-training productivity.[8] This negative effect cannot be accounted by human capital depreciation but can only be explained by the mechanism of feedback, in particular negative feedback.

In addition, receiving a bad grade can simply be discouraging for the participant as it reflects poor examination performance though the participant may still have made an effort. Some might even have studied hard and are frustrated about the fact that others received better grades and get, in addition, rewarded with the remission of fees. This is similar to the situation where students expect a reward for high effort - in the sense of a good grade - rather than high performance. Interestingly, the pay raise of 100 Euros for finishing the course is little-noticed. This is in line with behavioral evidence where a flat-rate premium that is received by all participants has no impact on the individual's performance. However, the negative outcome may signal reduced career opportunities.

We cannot answer the question whether high performers benefit from positive feedback because increased sales after training can be accounted for by both higher human capital as well as higher motivation through positive feedback. Because incentives from the company are identical for A- and B-graders, we would assume that both experience an increase in human capital from training. When looking at the progression of B-graders however, it can be argued that A-graders indeed build up human capital that is persistently used. Though B-graders finish final examinations with 81 – 92 % correct answers, they apply none of the information in any training phase, which leads to the suspicion that incentives are wrongly targeted on good grades as opposed to higher performance. As a result, B-graders can tab the refund without generating revenues. Abolishing the reward can be a way to improve results for the company. Because the reward does not increase a high performer's performance - sales of B-performers are stable and that of A-performers are already higher before examination - removing it would be less 'punishing' for low performers

[8] These effects stemming from skill attrition or lower motivation are sometimes found in trainings for unemployed (Merkl and Snower 2008).

who devote time and resources to training, receive negative feedback compared to others, and furthermore have to pay for the full training on their own.[9]

Conclusion

We investigate whether, how and when a company training program increases productivity at the workplace when taking individual learning outcomes and further unobserved heterogeneity into account. The program increased sales performance of participants who were later successful in the final exam *during* the training. However, this effect vanished at the end of the training, and there was no effect of positive feedback on performance. Moreover, for those participants who obtained a bad grade in the final exam, the training had no effect on performance during the training, but sales performance substantially *decreased* by approximately 6% below their prior level after the end of the training program. We found a significant negative effect of negative feedback on performance. Both results indicate that indirect (de-) motivating effects of training from feedback can be much more important for performance compared to the direct human capital formation that is supposed to be the aim of the training. Studying the timing of performance effects during the long-term e-Learning training shows that the performance effects are highest when preparing for the final examination. This suggests that the benefits of the final examination at the end of the training program need to be carefully traded against the negative effects of feedback.

[9] As managers of a franchise store should have a long-lasting financial incentive to keep up with high performance even after the training has finished, we re-estimated our model interacting training with being a manager of a franchise store but find no significant difference in this subgroup. However, this result has to be judged carefully due to limited amounts of franchise stores in our data set.

References

Agarwal, R., Angst, C.M., Magni, M., 2009. The Performance Effects of Coaching: A Multilevel Analysis using Hierarchical Linear Modeling. The International Journal of Human Resource Management 20, 2110–2134.

Alliger, G.M., Janak, E.A., 1989. Kirkpatrick's Levels of Training Criteria: Thirty years later. Personnel Psychology 42, 331–342.

Alliger, G.M., Tannenbaum, S.I., Jr, W.B., Traver, H., Shotland, A., 1997. A meta-analysis of the relations among Training Criteria. Personnel Psychology 50, 341–358.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: an Empiricist's Companion. Princeton University Press.

Anseel, F., Lievens, F., Schollaert, E., 2009. Reflection as a Strategy to enhance Task Performance after Feedback. Organizational Behavior and Human Decision Processes 110, 23–35.

Arthur, W., Bennett, W., Edens, P.S., Bell, S.T., 2003. Effectiveness of Training in Organizations: A meta-analysis of Design and Evaluation Features. Journal of Applied Psychology 88, 234–144.

Barrett, A., O'Connell, P.J., 2001. Does Training Generally Work? The Returns to In-Company Training. Industrial and Labor Relations Review 54(3), 647–662.

Barron, J.M., Berger, M.C., Black, D.A., 1998. Do Workers Pay for On-the-Job-Training. The Journal of Human Resources XXXIV 2, 235–252.

Bartel, A., 1989. Formal Employee Training Programs and Their Impact on Labor Productivity: Evidence from a Human Resources Survey.

Bartel, A., 1995. Training, Wage Growth, and Job Performance: Evidence form a Company Database. Journal of Labor Economics 13(3), 401–425.

Becker, G., 1964. Human Capital. Columbia University Press, New York.

Bishop, J., 1990. Job Performance, Turnover, and Wage Growth. Journal of Labor Economics 8(3), 363–386.

Björkman, I., Ehrnrooth, M., Mäkelä, K., Smale, A., Sumelius, J., 2013. Talent or Not? Employee Reactions to Talent Identification. Human Resource Management 52, 195–214.

Black, S.E., Lynch, L.M., 1996. Human-Capital Investments and Productivity. The American Economic Review 86(2), 263–267.

Dearden, L., Reed, H., VanReenen, J., 2006. The Impact of Training on Productivity and Wages: Evidence from British Panel Data. Oxford Bulletin of Economics and Statistics 68(4), 397–421.

Delaney, J.T., Huselid, M.A., 1996. The Impact of Human Resource Management Practices on Perceptions of Organizational Performance. Academy of Management Journal 39(4), 949–969.

Drukker, D.M., 2003. Testing for Serial Correlation in Linear Panel-data Models. The Stata Journal 3, 168–177.

Goodman, J.S., Wood, R.E., Chen, Z., 2011. Feedback specificity, information processing, and transfer of training. Organizational Behavior and Human Decision Processes 115, 253–267.

Gray, D.E., 2006. Executive Coaching: Towards a Dynamic Alliance of Psychotherapy and Transformative Learning Processes. Management Learning 37, 475–497.

Hanssens, D.M., Parsons, L.J., Schultz, R.L., 2003. Market Response Models - Econometric and Time Series Analysis, 2nd ed. Kluwer Academic Publishers.

Hinerasky, C., Fahr, R., 2014. When the Early Bird Catches the Worm: The Impact of training in retail. Working Paper, University Paderborn.
Holzer, H.J., Block, R.N., Cheatham, M., Knott, J.H., 1993. Are Training Subsidies for Firms Effective? The Michigan Experience. Industrial and Labor Relations Review 46(4), 625–636.
Kirkpatrick, D.L., 1979. Techniques for Evaluating Training Programs. Training & Development Journal 33, 78.
Kirkpatrick, D.L., Kirkpatrick, J.D., 2006. Evaluating Training Programs, 3rd ed. Berrett-Koehler.
Kluger, A.N., DeNisi, A., 1996. The Effects of Feedback Interventions on Performance: A historical Review, a Meta-analysis, and a Preliminary Feedback Intervention Theory. Psychological Bulletin 119, 254–285.
Kopelman, R.E., 1982. Improving Productivity through Objective Feedback. A Review of the Evidence. National Productivity Review 2, 43–55.
Kopelman, R.E., 1986. Objective Feedback. In: Edwin A. Locke, Generalizing from Laboratory to Field Settings, Lexington Books 119–145.
Kraiger, K., Ford, J.K., Salas, E., 1993. Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. Journal of Applied Psychology 78, 311–328.
Leivo, A.K., 2001. A field Study of the Effects of gradually terminated Public Feedback on Housekeeping Performances. Journal of Applied Social Psychology 31, 1184–1203.
Liu, X., Batt, R., 2007. Economic Pay-Offs to Informal Training: Evidence from Routine Service Work. Industrial and Labor Relations Review 61(1), 75–89.
Liu, X., Batt, R., 2010. How Supervisors Influence performance: a multilevel study of Coaching and Group Management in Technology-mediated Services. Personnel Psychology 63, 265–298.
Locke, E.A., Saari, L.M., Shaw, K.N., Latham, G.P., 1981. Goal Setting and Task Performance: 1969-1980. Psychological Bulletin 90, 125–152.
Ludwig, T.D., Biggs, J., Wagner, S., Geller, E.S., 2001. Using Public Feedback and Competitive Rewards to Increase the Safe Driving of Pizza Deliverers. Journal of Organizational Behavior Management 21, 75–104.
Magazine, T., 2010. 2010 Training Industry Report, pp. 18-31.
Mathieu, J.E., Tannenbaum, S.I., Salas, E., 1992. Influences of Individual and Situational Characteristics on Measures of Training Effectiveness. The Academy of Management Journal 35, 828–847.
McFarland, C., Miller, D.T., 1994. The Framing of Relative Performance Feedback: seeing the Glass as half empty or half full. Journal of Personality and Social Psychology 66, 1061–1073.
Merkl, C., Snower, D.J., 2008. Escaping the Unemployment Trap: The Case of East Germany. Journal of Comparative Economics 36, 542–556.
Montague, W.E., Webber, C.E., 1965. Effects of Knowledge of Results and Differential Monetary Reward on sic uninterrupted hours of Monitoring. Human Factors 7, 173–180.
Morrow, C.C., Jarrett, M.Q., Rupinski, M.T., 1997. An Investigation of the Effect and Economic Utility of Corporate-Wide Training. Personnel Psychology 50, 91–110.
Noe, R.A., 1986. Trainees' Attributes and Attitudes: Neglected Influences on Training Effectiveness. Academy of Management Review 11, 736–749.
Noe, R.A., Schmitt, N., 1986. The Influence of Trainee Attitudes on Training Effectiveness: Test of a Model. Personnel Psychology 39, Personnel Psychology.

Nordstrom, R., Lorenzi, P., Hall, R.V., 1990. A Review of public posting of Performance Feedback in Work Settings. Journal of Organizational Behavior Management 11, 101–123.

Olivero, G., Bane, K.D., Kopelman, R.E., 1997. Executive Coaching as a Transfer of Training Tool: effects on Productivity in a public Agency. Public Personnel Management 26, 461–469.

Richard, E.M., Diefendorff, J.M., 2011. Self-regulation during a single performance episode: Mood-as-information in the absence of formal feedback. Organizational Behavior and Human Decision Processes 115, 99–110.

Runnion, A., Watson, J.O., McWhorter, J., 1978. Energy Savings in Interstate Transportation Through Feedback and Reinforcement. Journal of Organizational Behavior Management 1, 180–191.

Sitzmann, T., Johnson, S.K., 2012. When is ignorance bliss? The effects of inaccurate self-assessments of knowledge on learning and attrition. Organizational Behavior and Human Decision Processes 117, 192–207.

Stephens, S.D., Ludwig, T.D., 2005. Improving Anaesthesia Nurse Complicance with Universal Precautions Using Group Goals and Public Feedback. Journal of Organizational Behavior Management 25, 37–71.

Sugrue, B., Rivera, R., 2005. State of the Industry. American Society for Training and Development's Annual Review of Trends in the Workplace, Learning and Performance. VA: American Society for Training and Development.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA. MIT Press.

Appendix

Table A1: Cross-tabulation Final Grade and Quartile

	Quartile				
Grade	1	2	3	4	Total
1 = A	768	0	0	0	768
2 = B	1,200	2,016	288	0	3,504
3 = C	0	0	1,632	1,296	2,928
4 = D	0	0	0	672	672
5 = E	-	-	-	-	-
6 = F	0	0	0	48	48
Total	1968	2,016	1,920	2,016	7,920

Note: Calculation of Quartiles is based on the percentile in the final exam.

Table A2: Average Final Grades and Points by Cohort

Cohort		**Written Test**		**Oral Test**		**Final Points**	**Final Grade**
	Freq.	Points	Grade	Points	Grade		
1	912	85	1.95	77	2.53	83	2.16
2	912	83	2.26	81	2.37	82	2.21
3	528	81	2.45	73	2.82	79	2.64
4	672	85	2.00	78	2.29	83	2.36
5	912	83	2.11	77	2.53	81	2.47
6	768	79	2.88	78	2.44	79	2.69
7	768	85	2.19	78	2.44	82	2.44
8	576	77	2.67	72	2.92	76	2.75
9	1,104	82	2.43	83	2.13	82	2.39
10	768	78	2.69	67	3.13	75	2.81

Note: Grades are given on a standardized basis: 100-92 Points Grade 1(A); <92-81 Points Grade 2(B); <81-67 Points Grade 3(C); <67-50 Points Grade 4(C); <50-30 Points Grade 5(E); <30-0 Points Grade 6(F).