Commonwealth Research Consulting, Inc.
Bulletin

CwRC–NIC Bulletin 5
March 2009

National Institute of Corrections Training Evaluation Project

2008 Evaluation Supplement: Learning, Application, and Action Plan Progress

By James B. Wells, Ph.D.
Kevin I. Minor, Ph.D.
and J. Stephen Parson

This is the fifth in a series of research bulletins on NIC's Training Evaluation Project. The project is being conducted by a team of researchers from Commonwealth Research Consulting, Inc. (CwRC), in collaboration with NIC's Division of Research and Evaluation. The purpose of the project, and the bulletin series, is to enhance understanding of training programs, and when appropriate, facilitate program improvements to better serve the field.

Highlights

- ***Data:*** 712 surveys[1] completed by 84 training participants and 11 trainers were analyzed for the bulletin. Response rates averaged 99% on initial surveys and over 90% on follow-up surveys (Table 1).
- ***Multiple Regression*** results were consistent with previous findings[2] described in Bulletin 3:
  - ⇒ Pre-training Model (Personal and Demographic variables) was a significant predictor for about 17% of tested outcome variables (1 of 6).
  - ⇒ Training Model (Quality, Type, and Relevance Measures) was a significant predictor for 50% of tested outcome variables (3 of 6).
  - ⇒ Post-training Model (Organizational Measures) was a significant predictor for 75% of tested outcome variables (3 of 4).
- ***Learning (post measure):*** On average participants reported a high to very high level of training-related learning at the close of training (Figures 1a & 1b). The training model was the best predictor of learning (Figure 1c).
- ***Learning (pre-post change):*** On average participants reported moderate to large learning gains during training (Figures 2a & 2b). The training model was the best predictor of learning gain scores (Figure 2c).
- ***Application of Learning:*** On average participants reported extensive post-training application of learning (Figures 3a & 3b). Both the training model and the post-training model were significant predictors of future application of learning (Figure 3c).
- ***Application Estimates:*** 50% of participants over-estimated the future application of training-related learning to their jobs, while less than 15% exceeded estimates (Figures 4a & 4b). The pre-training model was the best predictor, with older, more educated, and ***less*** experienced participants generally more accurate in their estimates (Figure 4c; pgs 7-9).
- ***Action Plan Progress:*** Participants on average reported moderate progress on action plan goals after training (Figures 5a & 5b). The post-training model (particularly funding, infrastructure, and support from management) was the best predictor of action plan progress (Figure 5c).
- ***Progress Estimates:*** Over 75% of participants fell short of estimated progress on action plan goals; 40% fell far short (Figures 6a & 6b). The post-training model was the best predictor estimate accuracy (Figure 6c).
- ***Lack of Progress:*** Less than 14% of participants indicated that NIC could have done anything differently to assist them where they fell short of estimated progress on action plan goals. Almost 40% indicated that they themselves could have done something to improve their progress. Over 75% indicated that their agency could have better supported their progress (Table 5).
- ***Bias and Discrimination:*** Analyses revealed no evidence that learning, application of learning, or action plan progress was significantly influenced by bias or discrimination based on age, race, gender, or any other demographic variable tested.

Training Evaluation Project Primary Staff

For Commonwealth Research Consulting, Inc:

James B. Wells, Ph.D.
President and Chief Research Consultant
jbwells@cwrc.us

Kevin I. Minor, Ph.D.
Senior Research Consultant
kiminor@cwrc.us

J. Stephen Parson
Research Consultant
jsparson@cwrc.us

For the National Institute of Corrections:

Christopher A. Innes, Ph.D.
Chief, Research and Evaluation
cinnes@bop.gov

Dee L. Halley
Correctional Program Specialist and Project Monitor
dhalley@bop.gov

Acknowledgements

The National Institute of Corrections Training Evaluation Project is made possible by the support of NIC via Cooperative Agreements 05A28GJF9, 06PEI01GJM1, and 07PEI12GJQ7.

CwRC staff wish to acknowledge the support and cooperation of the many persons who helped make this project possible. Morris Thigpen, Tom Beauclair, Chris Innes, Dee Halley, Virginia Hutchinson, Belinda Watson, Fran Zandi, Cheryl Paul, Robbye Braxton-Mintz, Rob Jeffreys and others at NIC have provided essential support for this project. We also wish to acknowledge our support staff, whose daily efforts further the project in so many ways. Finally, we want to express our appreciation to the growing number of NIC trainers and training participants who have taken time out of their busy schedules to graciously share their insights with us.

Although many persons and organizations contributed to the project described in this bulletin, any errors or omissions are those of the authors alone.

The findings, interpretations, and views presented in this bulletin are those of the authors and do not necessarily reflect the positions or policies of the National Institute of Corrections, or any other organization or individual.

Previous bulletins in the series include:

1. *Participant Demographics, Overall Evaluation of Training, and Applicability Ratings* (February 2007) provides a demographic sketch of 458 training participants, a discussion of early results from the evaluation project, and a preliminary profile of organizational resources and barriers to the implementation of training objectives in the workplace.
2. *Participant Evaluation of Trainers* (July 2007) focuses on 34 trainers involved in 20 Academy Division trainings offered during the pilot phase of the project (2005-2006), and provides a discussion of both quantitative and qualitative findings.
3. *Training Results, Activity Level Changes, and Implementation Results* (February 2008) discusses findings from a series of multivariate analyses of the relationships between demographic characteristics, training quality, post-training environments, and the successful implementation of training objectives in the organization.
4. *2008 Evaluation Results: Satisfaction, Learning, and Action Plan Progress* (November 2008) provides preliminary evaluation results from four FY08 Jails Division and Prison Division trainings.

These bulletins are available at: www.nicic.org/research.[3]

While earlier bulletins in the series focused on the evaluation of Academy Division trainings conducted during the pilot phase of the project, more recent bulletins focus on the evaluation of Jails Division and Prisons Division trainings. The FY08 training evaluations discussed in Bulletin 4 (2008 Evaluation Results) and Bulletin 5 (2008 Evaluation Supplement) include:

- Inmate Behavior Management (IBM: 08-J2301)
- Administering the Small Jail (ASJ: 08-J2801)
- Conducting Prison Security Audits (CPSA: 08-P3202)
- How to Run a Direct Supervision Housing Unit: Training for Trainers (DSHU: 08-J2202)

*The primary purpose of the current bulletin is to **update** Bulletin 4 findings to include recent CPSA and DSHU follow-up data, and to **expand** Bulletin 4 findings to include multivariate results based on all available data.*

As a supplement, rather than a replacement, this bulletin is best viewed together with Bulletin 4. The previous bulletin provides necessary background, and is referenced numerous times in the current bulletin. Note also that findings discussed in the current bulletin support important multivariate findings from Bulletin 3.

Data sources and response rates for the current bulletin are summarized in Table 1. Demographic data and initial evaluation data were collected from both participants

(N=84) and trainers (N=11). The decision to collect data from trainers as well as training participants arose from observations during the pilot phase that suggested an imbalance in the evaluation design. Collecting data on trainer demographics and trainer evaluation of participants, in addition to data collected from participants, facilitates the examination of issues such as potential age, race, or gender bias, and provides external or corroborative measures of participant attitude, learning, and participation levels. Evaluations based on more symmetric data provide a basis for the examination of previously inaccessible areas, for example, relationships between learning and trainer ratings of participant attention.

Initial evaluation data were collected in person via written forms. All forms included the necessary instructions and contact information for relevant NIC and CwRC personnel. A representative from NIC's Division of Research and Evaluation (R&E) was available at each training to introduce the evaluation, answer questions, and administer the necessary forms. Response rates were exceptionally high, averaging 99% (Table 1). This was due in large part to the personal involvement of R&E representatives during data collection, a revised evaluation protocol that included fewer and shorter surveys, and a combined information sheet and consent form that clearly communicated the purpose of the evaluation and the expectation of cooperation. Given that NIC paid all training-related expenses, including travel, accommodations, and per diem, and provided the trainings at no cost to participants, the research team concluded that language communicating an expectation of cooperation was appropriate. The previously used "voluntary" language was revised to read:

> *All trainers and training participants taking part in the [training name] program are expected to participate in the evaluation. Please give your full attention to each evaluation activity, as you would to any other aspect of the training program.*

Nonetheless, trainers and participants were informed they were free to decline to answer any question that made them uncomfortable, and that standard precautions would be taken to protect their privacy.

Follow-up evaluation data were collected from training participants beginning 90 days after conclusion of the training. Paper surveys were administered via US Mail. The follow-up data collection procedure was based on Dillman's Total Design Method (2000).[4] Although the multi-stage Total Design Method requires several months to execute, it is a well established method of maximizing survey response rates. Where the required four mailings fail to produce a completed follow-up, the current evaluation protocol provides for additional contacts by email and/or phone. Phone or email follow-up contacts by a person known to the potential respondent often overcome issues associated with outdated mailing addresses, procrastination, etc. For the current evaluation, these additional contacts were made by the CPS in charge of each program. These extra efforts increases the ASJ response rate from 72.7% to 100%, and the DSHU response rate from 70.8% to 83.3%. Note that in the case of IBM, additional contacts were not attempted with the three non-respondents because they were no longer employed by the sending agency. (The obtained response rate of 85.7% is nonetheless exceptionally good.) Likewise, in the case of CPSA, additional contacts were not attempted due to exceptional response to the standard mailings (94.4%).

Findings from a series of univariate and multivariate analyses of these data are provided next. The bulletin concludes with a summary and recommendations, followed by a discussion of the future directions of the evaluation project, and the bulletin series.

TABLE 1: Data Sources and Response Rates, 2008

Training Information					Completed Evaluation Forms (N=712)							Response Rates	
Training abbreviation, Code, and Dates; Number of Participants and Trainers					Participant Demographics	Trainer Demographics	Participant Evaluation of Training	Participant Evaluation of Trainers	Trainer Evaluation of Class	Trainer Evaluation of Participants	Participant Action Plan Follow-up	Initial	Follow-up
IBM	08-J2301	12/2 - 12/8/07	21	4	21	4	21	80	3	81	18 [a]	97.7%	85.7%
ASJ	08-J2801	1/13 - 1/18/08	22	3	22	3	22	66	3	65	22	99.1%	100.0%
CPSA	08-P3202	6/2 - 6/6/08	18	3	18	3	18	36	3	18 [b]	17	100.0%	94.4%
DSHU	08-J2202	6/9 - 6/19/08	24	2	24	2	24	48	2	48	20	100.0%	83.3%
TOTALS			**85** [c]	**12** [d]	**85** [c]	**12** [d]	**85** [e]	**230** [e]	**11** [f]	**212** [f]	**77**	**99.0%**	**90.6%**

[a] Three Inmate Behavior Management participants were no longer employed by the sending agency at the time of the 90 day action plan follow-up.

[b] CPSA training was split into three groups (1 trainer and 6 participants each) for most of the training; each trainer evaluated only those 6 participants.

[c] There were 84 unique participants (one participant attended two trainings.)

[d] There were 11 unique trainers (one trainer was used in two trainings.)

[e] A condensed, combined form was used to allow participants to evaluate the training overall, and each trainer individually.

[f] A condensed, combined form was used to allow trainers to evaluate participants individually, and as a group.

Findings

Note that findings from Bulletin 4[5] have been updated to include recent CPSA and DSHU follow-up data. Updated findings are included here where appropriate, though most appear under Action Plan Progress, and in Tables 3, 4, and 5. For demographic profiles of participants and trainers, satisfaction ratings, and other 2008 evaluation results not included in the current bulletin, see Bulletin 4.

Findings from the analyses of initial data and follow-up data provided by 84 training participants and 11 trainers in four 2008 trainings (Table 1) are presented in three sections to follow. The first section, ***Learning,*** provides evaluation results regarding changes in participants' training-related knowledge, skills, and positive attitudes during training. The next section, ***Application of Learning*****,** examines the extent to which participants applied the training to their jobs. The final section, ***Action Plan Progress,*** examines the extent to which training-related learning, and the application of learning, translate into desired outcomes in the organization, i.e., progress on action plan goals after training.

In each of these areas, analyses revealed considerable variation in participant outcomes. Therefore, ***potential sources of outcome variation*** were examined in greater detail. Several hundred variables from the seven surveys listed in Table 1 were considered as possible outcome predictors. Based on previous research, an evolving theory of training outcomes, and results from preliminary analyses, approximately 45 of those variables were selected and arranged into three sets or models:[6]

1. ***Pre-training measures*** selected for the model were age, race, gender, education, years on current job, total experience, and salary.
2. ***Training measures*** included several variables of training quality, training relevance, trainer effectiveness, training type, etc. In addition to participant ratings of trainers and training, trainer ratings of participants were also included.[7, 5]
3. ***Post-training measures*** consisted primarily of organizational variables, such as various resources and barriers to implementing training objectives on the job, but also include participants' post-training effort levels, post-training contact with trainers, etc.

Multiple regression modeling was then used to test the predictive or explanatory power of the three models. Although multiple regression modeling is an advanced statistical technique, the basics of the procedure are relatively simple: it tests the predictive power of each variable in the model, independently and collectively. In lieu of further explanation or lengthy examples, suffice it to say that ***no technical understanding of the procedure is required*** to grasp findings as presented in the bulletin narrative. Although multiple regression modeling produces several statistics, only the percentage of explained outcome variation[8] is presented in the bulletin narrative; some of the more technical results are presented in endnotes.[9]

The extent to which each of the three regression models predict or explain variation in ***Learning*****,** ***Application of Learning,*** or ***Action Plan Progress*****,** is discussed in each of those sections to follow. When interpreting the percentage of outcome variation explained by each of the three models, please note the following:

1. ***Pre-training measures*** such as age, race, or gender are expected to explain very little outcome variation in relatively uniform or homogenous settings such as training environments, unless there is bias in the selection or treatment of training participants. In more stratified environments, such as the realities of the workplace in correctional organizations, demographic variables may explain somewhat more outcome variation simply because some persons in the organization are more highly placed, or more empowered to act, than others. In either environment, however, pre-training measures are expected to explain relatively little outcome variation compared to *training measures* or *post-training measures*.
2. ***Training measures*** typically explain relatively more outcome variation, especially early and intermediate outcomes, because, theoretically and practically, one would expect variables such as training quality and relevance to be associated with (and predictive of) positive outcomes.
3. ***Post-training measures*** typically explain more outcome variation than either of the other two models, especially with regard to distal or ultimate outcomes. This is neither good nor bad, per se. It simply reflects the common sense (and research supported) notion that the organization matters, and that training graduates must contend with the complexities and difficulties of post-training environments in order to successfully transfer training substance to the workplace and achieve action plan goals after training.
4. Although ***bivariate correlations*** are discussed throughout the bulletin series, multivariate results, such as those derived from multiple regression modeling, should be given precedence when results appear to conflict. Although a full explanation is beyond the scope of this bulletin, in general this is because properly executed multivariate techniques take into account a wide variety of potentially confounding influences, while bivariate techniques are far more limited.

Learning

In this section, learning is examined both from the perspective of post-training levels, and the difference between pre-training and post-training levels, i.e., pre-post change scores. The data were collected as part of a participant self-assessment of learning conducted at the conclusion of each training.[10] Participants were asked to rate their level of learning, prior to and after training, in each of three areas: knowledge, skills, and positive attitudes regarding training topics. Participants were also asked to rate the extent to which they had been able to apply each of these prior to training, and the extent to which they anticipated being able to apply them after training. (See Bulletin 4 for detailed descriptive results for each individual area.) The five point rating scale ranged from very low (1) to very high (5).

Learning: Post-training Measures

As illustrated in Figure 1a, participants of each training reported high to very high levels of post-training learning on average, with an overall mean of 4.55 (N=84, SD=0.44). CPSA and DSHU participants reported slightly higher levels of post-training knowledge, skills, and positive attitudes overall than did participants of IBM and ASJ. Nonetheless, these differences are small and probably of no practical significance.

As illustrated in Figure 1b, the distribution of learning scores for the 84 participants was somewhat narrow, (SD=0.44) ranging from means of 3.50 (medium-high) to 5.00 (very high). Nonetheless, this range of outcomes is sufficiently broad to warrant evaluation of potential sources of the variation. As illustrated in Figure 1c, measures of training quality and relevance accounted for 16.1% of the variation[11] in reported post-training levels of learning, i.e., participants who gave higher ratings to training quality were more likely to report higher levels of training-related knowledge, skills, and positive attitudes after training.

Although the training measures model does not appear to be a particularly strong predictor of learning, this is due in part to the relatively narrow range of variation (Figure 1b) in reported learning, i.e., the less variation there is, the less accurately any model can explain or predict the variation. A more useful way to gauge the strength or utility of the training measures model, is to compare its performance against that of the other models tested. The pre-training model, for example, was not significant,[12] i.e., none of the variables included in this model, individually or in combination, reliably accounted for any significant variation in reported post-training learning levels. Thus, measures of training quality and relevance performed better than pre-training or personal measures in terms of predicting or explaining variation in reported learning.

Both results are quite positive. Although 16.1% explained variation in learning is a relatively weak result for the training quality model, it nonetheless indicates a clear connection between training quality and learning. On the other hand, and perhaps more important, the lack of significant results from the pre-training model provide no evidence of bias, discrimination, or differences in learning based on age, race, gender, or any other demographic or personal variable included in the model. In the context of a training environment, this suggests fair and unbiased selection and treatment of training participants.

Note that the post-training model was not tested as potential predictors of learning because of the obvious conflict in temporal order, i.e., the training-related learning in question occurred prior to post-training factors coming into play.

Learning: Pre-training to Post-training Change Scores

As illustrated in Figure 2a, participants of each training reported mean learning gains in the neighborhood of one to two units, with an overall mean of 1.49 (N=84, SD=0.78). For example, a pre-to-post change from low to medium (2.0 to 3.0), or medium to high (3.0 to 4.0) would represent a gain of one unit, while a change from low to high (2.0 to 4.0) or medium to very high (3.0 to 5.0) would represent a gain of two units. Participants of CPSA reported somewhat lower gains on average than participants of the other trainings (0.97 versus 1.44 to 1.76). As discussed previously in Bulletin 4 (page 11) this is due primarily to the higher pre-training levels of learning reported by the typical CPSA participant (3.66), relative to participants of the other trainings (2.79 to 2.94). Note also that the demographic

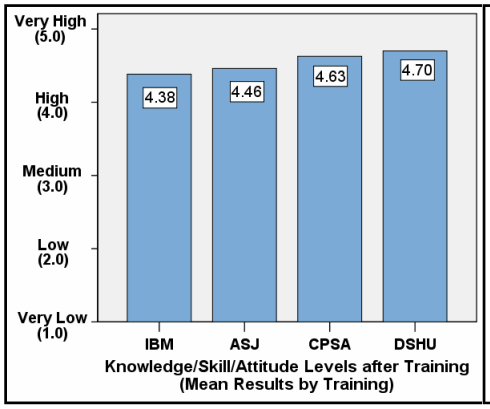

Figure 1a

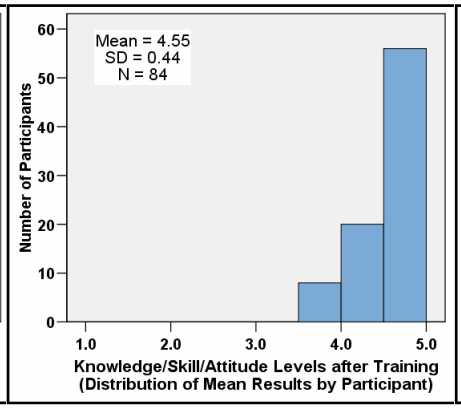

Figure 1b

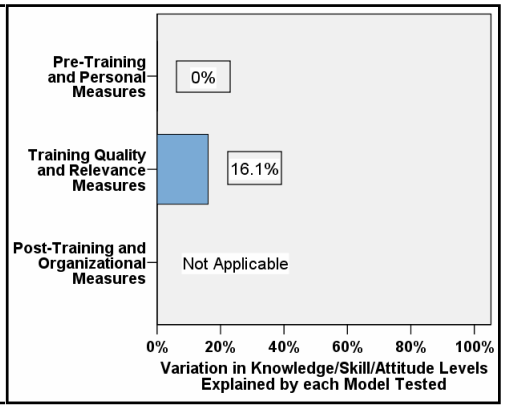

Figure 1c

profile of CPSA participants was consistent with the higher reported pre-training knowledge/skill levels, i.e., they were more educated, more experienced, and more advanced in their careers, on average. This higher pre-training or baseline level, in conjunction with similar post-training levels (Figure 1a), accounts for the somewhat lower pre-to-post gains reported by CPSA participants relative to participants of the other trainings. However, despite the fact that evidence suggests CPSA participants "knew more going into the training" they nonetheless expressed significant pre-to-post learning gains, and provided high training satisfaction ratings (see Bulletin 4).

As illustrated in Figure 2b, the distribution of learning change scores for the 84 participants is quite broad (SD=0.78) ranging from reported gains of zero (N=2)[13] to gains in excess of three units (N=3). Multiple regression modeling was used to evaluate potential sources of this outcome variation. As illustrated in Figure 2c, measures of training quality and relevance accounted for 28.4% of the variation[14] in reported learning gains. In this case, participants who rated their trainers higher also typically reported higher pre-to-post changes in learning. It is also important to note that despite somewhat lower learning gains reported by CPSA participants, training type, e.g., CPSA, IBM, etc., was ***not*** a significant predictor of learning gain. In other words, while statistically controlling for other variables in the model, trainer ratings remained a significant predictor of learning gains, but training type did not. Although this may seem counter intuitive, it is actually a reasonable and logical finding. For example, suppose a strong negative correlation between height and hair length was observed in a group of people, i.e., taller people generally had shorter hair. But after controlling for the effects of gender (men tend to be both taller and have shorter hair than women) the correlation between height and hair length disappears. Similarly, while controlling for the effects of other variables in the model, the relationship between training type and learning gain becomes insignificant, while the relationship between trainer ratings and learning gain persists.

The pre-training model, once again, was not significant, i.e., none of the demographic and personal variables included in this model, individually or in combination, were reliable predictors of learning gains.[15] Despite previously mentioned differences in the demographic profile of CPSA participants relative to participants of other trainings, when controlling for the effects of other demographic/personal variables in the model, no individual variables were significant, nor was the overall model. As mentioned previously, the lack of significant results from the pre-training model provides no evidence of bias, discrimination, or differences in learning based on age, race, gender, or any other demographic or personal variable included in the model. Obviously, in the case of the pre-training model, a finding of 0% explained variation in learning gain is a very positive finding.

Note that the post-training model was not tested as potential predictors of pre-training to post-training learning gains because of the obvious conflict in temporal order, i.e., the training-related learning in question occurred prior to post-training factors coming into play.

Application of Learning

Given the tightening budgets, limited staff, and challenging environments characteristic of corrections, it is crucial that all the goals of training are achieved. Obviously it is important that training participants are satisfied with the quality of training and adequately learn its content. However, training satisfaction and learning matter little without the successful transfer of learning to the workplace. For the last several decades researchers have estimated that staff resistance and other organizational factors typically block up to 90% of training substance from transferring to the workplace.[16] Likewise, although the evaluation of 20 NIC trainings offered during 2005 and 2006 found that measures of training quality and relevance were moderate predictors of learning transfer, organizational measures were the strongest predictors.[2] Thus measures such as organizational resources and barriers are especially relevant to the application of training-related knowledge,

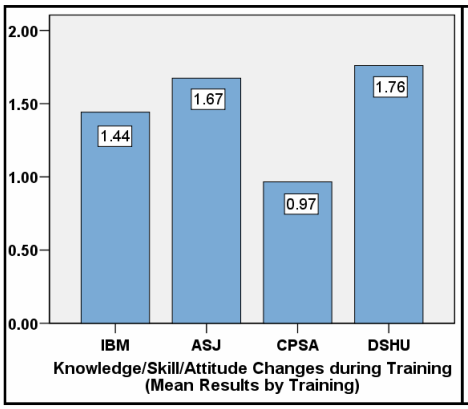

Figure 2a

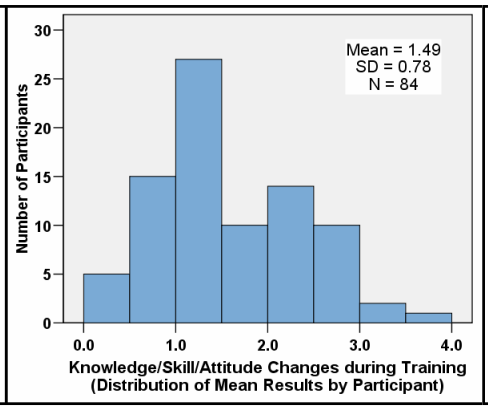

Figure 2b

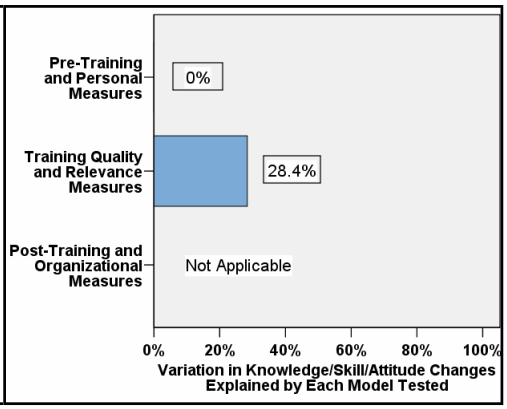

Figure 2c

skills, and attitudes in the participant's organization after training.

In this section, the extent to which participants applied training-related learning to their jobs after training is evaluated both from the perspective of post-training application levels, and the difference between application estimates made at the conclusion of training and 90 day follow-up reports of application, i.e., estimate accuracy. Initial data were collected as part of an overall training evaluation conducted at the conclusion of each training. Participants were asked to rate "the extent to which you anticipate applying what you learned in this training to your job." Follow-up data were collected via written surveys mailed 90 days after training. On the follow-up, participants were asked to rate "the extent to which you were able to apply the knowledge/skills gained from this training to your job responsibilities." On both forms, participants were asked to select a number from 0 (not at all) to 10 (a great deal). (See Bulletin 4 for additional descriptive and bivariate results.)[5]

Application of Learning to the Job

As illustrated in Figure 3a, participants of all trainings reported extensive application of training-related learning to their jobs in the three months following conclusion of the training. The overall mean was 8.15 (N=60, SD=2.10), and mean ratings for each training ranged from a low of 7.20 for IBM participants to a high of 8.94 for DSHU participants. Even the lowest of these mean ratings is quite high, and should be interpreted as a very positive finding.

As illustrated in Figure 3b, the distribution of application ratings for the 60 participants who provided follow-up data, though positively skewed, is quite broad (SD=2.10) ranging from 0 to 10. Though most ratings range from 6 to 10, this is more than sufficient variation to warrant evaluation of potential sources of the variation. In fact, as illustrated in Figure 3c, two of the three regression models were significant predictors of outcome variation.

The training quality model accounted for 27.3% of the variation[17] in reported application of training-related learning in the workplace, i.e., participants who gave higher ratings to training quality and relevance at the conclusion of training were more likely to report more extensive application of training content to their jobs 90 days after training.

The model composed of post-training and organizational measures was also significant, explaining 23.2% of the variation[18] in the application of training-related learning. This model consisted primarily of various post-training factors that training participants had previously identified as resources or barriers in the organization that helped or hindered them in applying what they learned in training to their jobs (Table 3). However, the model also included a variable for participants' self-assessment of the extent to which they themselves were either a resource or barrier in the application of training-related learning. Not surprisingly, participants who judged themselves to be resources, and who reported a more favorable balance of resources and barriers in their organization, were more likely to report more extensive application of training-related learning to their jobs. (See also Bulletin 4, pages 12-19.)[5]

Once again the pre-training model was not significant, providing no evidence of bias, discrimination, or differences in the application of learning based on age, race, gender, or any other demographic or personal variable included in the model.[19] In the context of post-training application of learning, this suggests participant motivation did not vary significantly based on age, race, gender, etc.

Estimating the Future Application of Learning

Sufficient mastery of training content, and the transfer of learning to the workplace are important goals of training. Often overlooked, however, is the broad range of organizational factors that can interact with, complicate, and impede the implementation of training objectives in the workplace. Moreover, some evidence suggests that the ability to accurately estimate the extent to which training objectives can realistically be implemented in a given amount of time, can in itself, facilitate progress on action plan goals and other desired outcomes. Conversely, evaluation results suggest that the inability to accurately

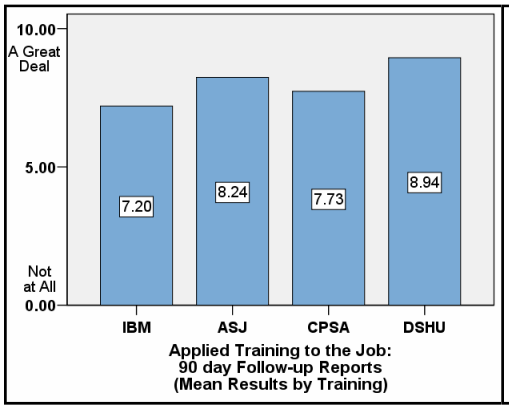

Figure 3a

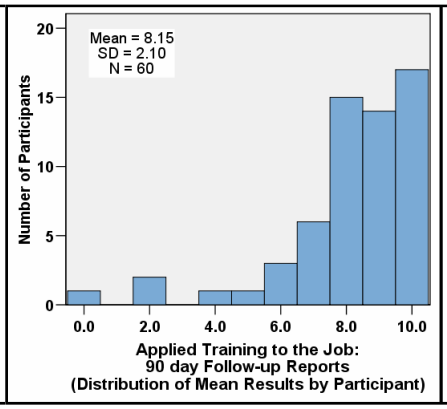

Figure 3b

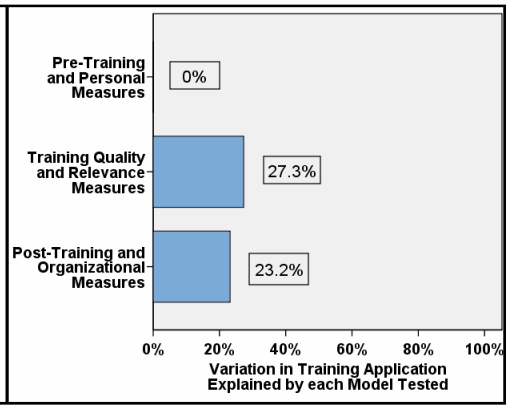

Figure 3c

estimate the future application of training-related learning can undermine progress on action plan goals and other desired outcomes.

As illustrated in Figure 4a, analyses of data from 90 day follow-up reports indicate that, on average, participants of all trainings significantly overestimated the extent to which they would apply training-related learning in their daily work.[20] The mean overestimation was –0.77 (N=48, SD=1.78). Means for each training ranged from –0.47 for participants of DSHU to –1.17 for IBM participants.

As illustrated in Figure 4b, mean differences between estimated application of learning and follow-up reports of applied learning for individual participants ranged from 3.0 to –7.0. However, mean differences for most participants ranged from 2.0 to –4.0. Overall, 50% of participants overestimated the extent to which they would apply training-related learning to their jobs, and participants who fell short of estimates outnumbered those who exceeded estimates by a margin of over three to one.

This pattern of overestimation appears to be unrelated to training quality and relevance. Multiple regression results indicate that the training quality model was not a significant predictor of variation in the accuracy of training application estimates (Figure 4c).[21] Likewise, post-training measures such as organizational resources and barriers also failed to predict the accuracy of application estimates, despite testing several versions of the model.[22]

Pre-training and personal measures, on the other hand, proved to be very strong predictors of the extent to which participants accurately estimated the future application of training-related learning in their daily work (Figure 4c).[23] This model reliably accounted for over 55% of the variation in the accuracy of application estimates. Older and more educated participants were generally more accurate in their estimates than younger and less educated participants. On the other hand, participants with ***less*** time on the current job, and ***less*** total experience were also generally more accurate in their application estimates. While these findings may seem contradictory or counterintuitive, bear in mind that multiple regression modeling isolates the unique contribution of each predictor variable, while statistically controlling for the effects of all other predictor variables in the model. Thus it is entirely plausible that ***both*** older participants, ***and*** less experienced participants can be more accurate in their estimates, despite the fact that older participants were generally more experienced (r=.510, N=42, p.001) This is because regression modeling isolated the unique contribution of each of these variables while holding constant the effects of all other predictor variables in the model. For example, less experience in corrections, or in a particular job, (for participants of similar age) may predispose training participants to be relatively less confident and/or more conservative in their estimates of the extent to which they will apply training-related learning to their jobs. Crosstab results support this interpretation. To illustrate, consider the estimate accuracy ratings of the four groups of participants shown in Table 2.

Clearly, older participants were more accurate in their estimates, when controlling for experience (horizontal arrows): –.0.23 vs –3.67 and 0.50 vs –0.75. Likewise, less experienced participants were more accurate in their estimates, when controlling for age (vertical arrows): 0.50 vs –0.23 and –0.75 vs –3.67.

Table 2: Accuracy of learning application estimates, by age and experience

	Age 40+	Age <40
Experience 10+	-0.23	-3.67
Experience <10	0.50	-0.75

In some cases a significant pre-training and personal measures model can signal bias, discrimination, or differential benefit based on age, race, gender or other independent (predictor) variables included in the model. This is somewhat more likely in the case of more traditional outcome variables such as satisfaction, learning, or action plan progress. However, in the case of more esoteric measures, such as the estimate accuracy variable, a significant pre-training model is of less concern. The

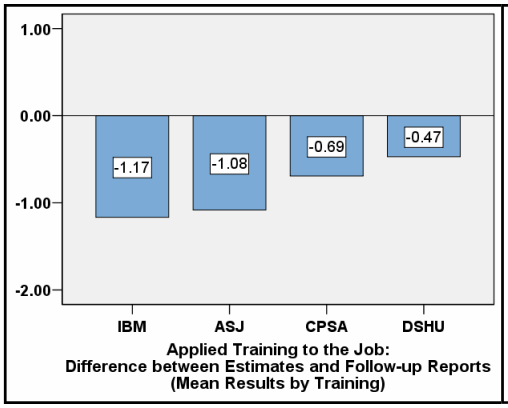

Figure 4a

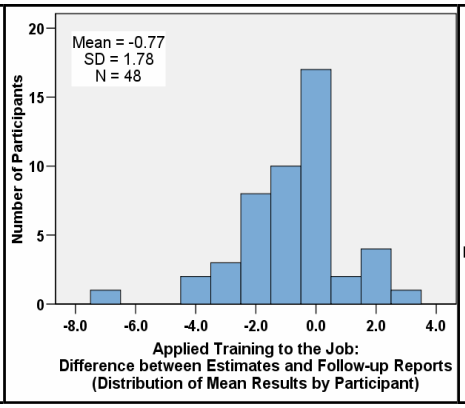

Figure 4b

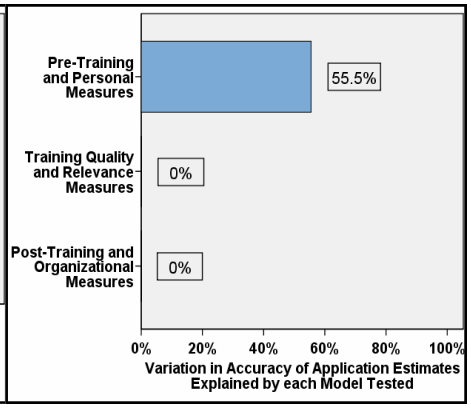

Figure 4c

identified relationships between estimate accuracy and age, education, and experience appear logical and reasonable, and should not be interpreted as indications of bias or discrimination.

In the next section, ***Action Plan Progress***, variables that have thus far been considered outcome measures will be entered into regression models and tested as potential predictors of more distant outcomes.

Action Plan Progress

Training-related learning, and the application of learning in the workplace, are important proximal or intermediate outcomes of training. Nonetheless, it is also important to evaluate the extent to which learning, and the application of learning, translate into desired outcomes in the organization, i.e., distal or ultimate outcomes. Training action plans can help participants organize their learning and focus their efforts on achieving tangible, measurable improvements in the organization. Progress on action plan goals is one way to evaluate distal outcomes.

In this section, action plan progress is evaluated both from the perspective of 90 day follow-up reports, and the accuracy of initial estimates done at the time of training. Initial data were collected from training action plans completed by each participant at the time of training. In each action plan, participants described one or more goals they intended to pursue, as a result of training, upon returning to their agency. Action plans also specified the amount of progress participants expected to make toward each goal in the 90 days following training, and how they intended to measure progress. Over 88% of training participants completed a training action plan, expecting to make moderate to substantial progress on their goals, on average, during the follow-up period (Mean=3.58, N=75, SD=0.72). Anticipated progress means for each training ranged from 3.37 for CPSA participants (N=17, SD=0.51) to 3.78 for DSHU participants (N=23, SD=0.85).

Action Plan Progress Follow-up Reports

As part of the 90 day action plan follow-up, participants were asked to indicate the amount of progress they had actually made on each goal. As illustrated in Figure 5a, ASJ participants reported moderate to substantial progress on action plan goals, while participants of the other trainings reported moderate progress on average. The overall mean progress rating was 3.05 (N=76, SD=1.00). These are very favorable findings in the sense that they provide some evidence that participants have been moderately successful translating training objectives into improvements in the organization.

Although participants reported moderate action plan progress on average, results for individual participants varied considerably. As illustrated in Figure 5b, participants ranged from no progress (1.0) to exceptional progress (5.0) on action plan goals. While 27 participants (35.5%) reported approximately moderate progress (2.5 to 3.5) toward achieving their goals, 21 participants (27.6%) reported slight to no progress (<2.5). The remaining 28 participants (36.8%) reported substantial to exceptional progress (>3.5).

Multiple regression modeling was conducted to examine potential sources of variation in action plan progress. As illustrated in Figure 5c, the only significant model was post-training and organizational factors, which accounted for 28.6% of the variation in action plan progress.[24] Participants who identified a more favorable balance of organizational resources and barriers, particularly with regard to funding, infrastructure, and support from management, generally reported greater progress on their action plan goals.

Interestingly, the training measures model (Figure 5c) was not a significant predictor of action plan progress.[25] Thus measures of training quality and relevance, training type, trainer ratings, etc. appear to be unrelated to action plan progress. These are not particularly negative findings. They simply reflect the commonsense (and research supported) notion that despite the importance of training quality to intermediate outcomes, when it comes to translating these

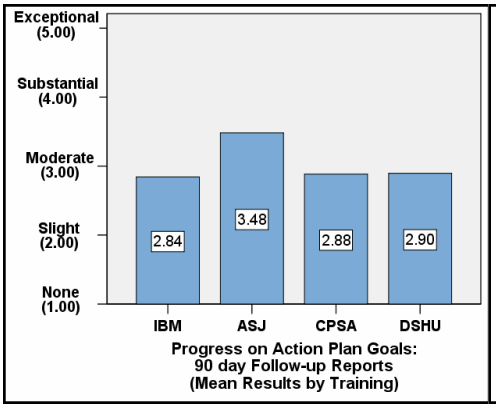

Figure 5a

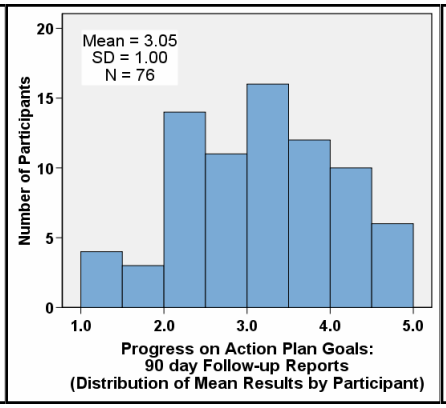

Figure 5b

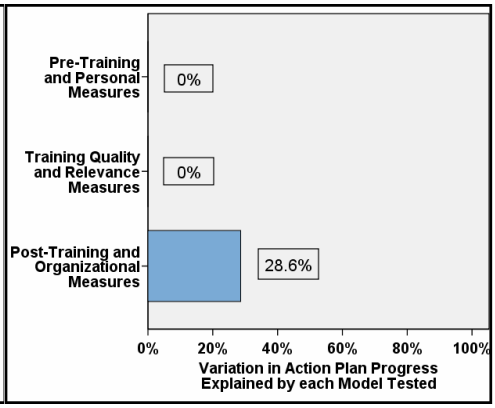

Figure 5c

into results in the organization, the organizational context overshadows training variables.

The model of pre-training and personal measures (Figure 5c) also failed to be a significant predictor of action plan progress.[26] As discussed previously, this is a very favorable finding in that it provides no evidence of bias, discrimination, or differences in action plan progress based on age, race, gender, or any other demographic or personal variable included in the model. In the context of post-training progress implementing action plans in the organization, this suggests that neither participant motivation and efficacy, nor organizational receptiveness varies significantly according to participant age, race, gender, etc.

Estimating Future Progress on Action Plan Goals

Accurately estimating future progress on action plan goals is important for a variety of reasons. For example, progress estimates may form the basis for commitments to or coordination with colleagues, superiors, other departments, or outside organizations. Progress estimates may also provide the basis for complying with current or emerging policies, standards, regulations, or laws. When actual progress falls short of estimates, reputation and feelings of efficacy may suffer, and relationships can become strained.

As illustrated in Figure 6a, analyses of data from 90 day follow-up reports indicate that, on average, training participants significantly overestimated the amount of progress they would make on action plan goals within 90 days after training.[27] The mean overestimation was –0.56 (N=67, SD=1.07). Means for each training ranged from –0.16 for ASJ participants, to –0.91 for DSHU participants.

As illustrated in Figure 6b, mean differences between estimated action plan progress and follow-up progress reports for individual participants ranged from 4.0 to –4.0. However, mean differences for most participants ranged from 2.0 to –2.0. Nonetheless, only 24% (16 of 67) of participants met or exceeded their estimated 90 day progress on action plan goals. Overall, 76% (51 of 67) of participants who provided action plan data fell short of estimated progress. About 40% fell far short of estimates, i.e., at least one full unit, such as the difference between substantial and moderate progress, or approximately double the average shortfall. Note that despite the small mean shortfall of –0.16 for ASJ participants, 70% (14 of 20) fell short of achieving estimated progress, and 20% fell far short.

The tendency to overestimate action plan progress appears to be unrelated to training quality and relevance. Multiple regression results indicate that the training quality model was not a significant predictor of estimate accuracy (Figure 6c).[28] Likewise, the model consisting of pre-training and personal measures also failed to predict accuracy in progress estimates.[29] As mentioned previously, this is generally a very favorable finding in the sense that it provides no evidence of bias, discrimination, or differential benefit from training based on age, race, gender, or other demographic variables included in the model.

However, in the case of estimate accuracy, this finding raises the question of why pre-training and personal measures were strongly predictive of accuracy in learning application estimates (Figure 4c), but not predictive of accuracy in action plan progress estimates. One explanation that is both commonsensical and consistent with the findings is that estimating the ***results of ones future efforts***, e.g., action plan progress, is a more complicated and uncertain process than simply estimating ones ***future efforts***, e.g., applying learning to the job.[30] The increased complexity and uncertainty is due in part to the likelihood that any errors in estimated ***effort*** (learning application) may be magnified as even larger errors in estimated ***results*** (action plan progress). For example, the mean shortfall in learning application estimates was –0.77 on a 0 to 10 scale, or 7.0% of the scale. However, the mean shortfall in action plan progress was –0.56 on a 1 to 5 scale, or 11.2% of the scale. Likewise, the increased complexity and uncertainty of estimating ***results*** compared to estimating ***efforts*** may also stem from the likelihood that ***results***, in addition to being contingent upon ***efforts***, are also somewhat more exposed to, or dependent upon, organizational resources and barriers, while ***efforts*** are more centered in the individual and somewhat less

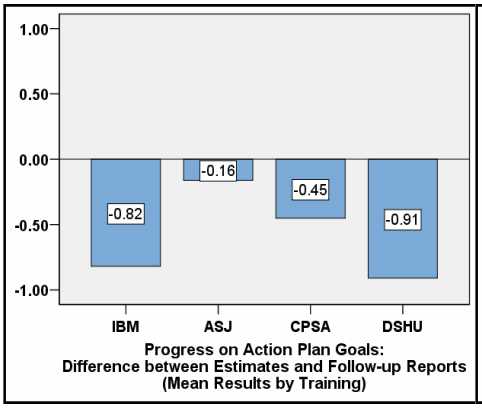

Figure 6a

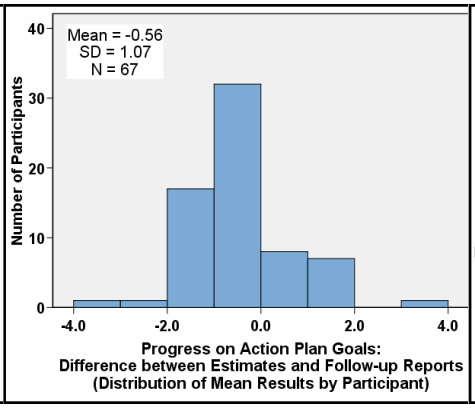

Figure 6b

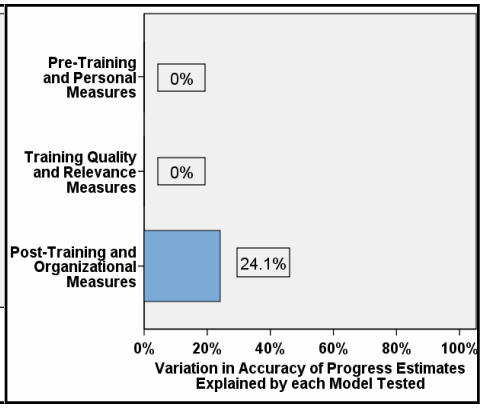

Figure 6c

dependent upon organizational factors. In short, the trainee has more control over effort than results.

This interpretation is supported by numerous findings from multiple regression modeling. As illustrated in Figure 6c, the post-training and organizational model is significant, and a better predictor of accuracy in action plan progress estimates than either of the other models. The model accounted for 24.1% of the variation in the accuracy of action plan progress estimates.[31] Specifically, participants who had more accurately estimated the extent to which they would apply training-related learning to their jobs, and who reported fewer organizational barriers and more resources, were generally more accurate in estimating the amount of progress they would achieve on action plan goals following training. Recall also that post-training and organizational factors were the best predictors of action plan progress (Figure 5c). Moreover, findings from Bulletin 3 demonstrate that organizational resources and barriers were consistently strong predictors of results in the form of implementation progress for both training-specific objectives, and participant-identified objectives.[2] Finally, as discussed in Bulletin 4, findings suggest that participants had considerable difficulty accurately estimating organizational resources and barriers.[5] This may undermine the utility of resources and/or increase the impact of barriers not only on accurately estimating efforts and their results, but on efforts and results themselves.

It is interesting to note that pre-training/personal measures were predictive of the accuracy of effort estimates (but not result estimates) while opposite findings were obtained for post-training/organizational measures, which were predictive of the accuracy of result estimates (but not of effort estimates). In short, available evidence suggests that:

__Effort__ is more contingent upon the __individual__, while __results__ are more contingent upon the __organization__.

Table 3: Participant Action Plan Follow-up **(Bulletin 4, Table 13 Update)**

(Survey items are paraphrased. Negatively worded items were recoded/rephrased for purposes of scaling and analysis.)	IBM			ASJ			CPSA			DSHU			Overall		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
90 Day Post-Training Evaluation															
Would recommend training to others with similar jobs	17	**4.35**	1.32	22	**4.27**	1.58	16	**4.19**	1.47	20	**4.25**	1.55	75	**4.27**	1.46
Training was relevant to my organization overall	17	**4.24**	1.30	22	**4.82**	0.39	16	**4.56**	1.03	20	**4.65**	0.49	75	**4.59**	0.86
Training was relevant to my job duties in particular	17	**4.24**	1.30	22	**4.86**	0.35	16	**4.44**	1.03	20	**4.50**	1.05	75	**4.53**	0.98
(...4 = Agree; 5 = Strongly Agree) **Overall**	17	**4.27**	0.99	22	**4.65**	0.60	16	**4.40**	0.78	20	**4.47**	0.82	75	**4.46**	0.80
Extent to which you were able to apply the knowledge/ skils gained from this training to your job: (0=Not at All; 10=A Greal Deal)	10	**7.20**	3.22	17	**8.24**	1.25	15	**7.73**	2.69	18	**8.94**	1.00	60	**8.15**	2.10
Encountered Resources and/or Barriers to Implementing Training in Participants' Agencies/Organizations															
Management/Administrative Support	18	**1.94**	1.66	22	**1.77**	1.38	17	**1.53**	1.37	20	**1.60**	1.70	77	**1.71**	1.51
You yourself	18	**1.72**	1.41	21	**1.86**	1.11	16	**1.25**	1.18	20	**1.70**	1.03	75	**1.65**	1.18
Support from a Key Person	18	**1.72**	1.78	22	**1.73**	1.39	17	**1.18**	1.51	20	**1.70**	1.17	77	**1.60**	1.45
Teamwork (Cooperation within your Agency/Org)	18	**1.00**	1.53	22	**1.59**	1.01	17	**1.06**	1.60	20	**1.15**	1.63	77	**1.22**	1.44
Staff Development/Training	18	**1.22**	1.44	22	**1.18**	1.65	17	**0.94**	1.48	20	**0.90**	1.86	77	**1.06**	1.60
Agency Structure/Policy	18	**0.17**	2.38	22	**1.45**	1.37	17	**1.12**	1.45	20	**0.70**	2.11	77	**0.88**	1.89
Existing Programs	18	**0.11**	1.64	22	**1.09**	1.23	17	**1.18**	1.24	20	**0.60**	1.60	77	**0.75**	1.47
Organizational Acceptance/Resistance	18	**0.11**	1.45	22	**1.05**	1.70	17	**1.12**	1.50	20	**0.70**	1.89	77	**0.75**	1.67
Cooperation between Agencies/Organizations	18	**-0.11**	1.91	22	**0.95**	1.59	17	**1.12**	1.58	20	**0.75**	1.55	77	**0.69**	1.69
Personnel (staffing levels, skills, experience, etc.)	18	**0.28**	1.90	22	**1.05**	1.86	17	**0.76**	1.79	19	**-0.32**	2.08	76	**0.46**	1.95
Other (please specify)	3	**0.00**	0.00	6	**0.83**	1.33	3	**0.00**	0.00	6	**-0.17**	1.60	18	**0.22**	1.22
Workload/Time	18	**0.11**	1.64	21	**0.10**	1.58	17	**0.00**	1.73	20	**0.20**	2.12	76	**0.11**	1.75
Funding/Infrastructure	18	**-0.83**	1.82	22	**0.64**	1.50	17	**-0.35**	1.77	20	**0.25**	1.62	77	**-0.03**	1.73
Overall	18	**0.60**	1.13	22	**1.19**	1.02	17	**0.91**	1.07	20	**0.81**	1.25	77	**0.89**	1.12
Mean progress on participant-identified goals since training: (1=None; 2=Slight; 3=Moderate; 4=Substantial; 5=Exceptional)	18	**2.84**	0.74	22	**3.48**	0.60	16	**2.88**	1.19	20	**2.90**	1.28	76	**3.05**	1.00

Updated Action Plan Progress Results from Bulletin 4

As previously noted, action plan follow-up data for CPSA and DSHU were not available at the time Bulletin 4 was published (November 2008). Follow-up data collection for those trainings is now complete. Updated response rates and data sources are summarized in Table 1 (page 3). Updated univariate (descriptive) findings are summarized in Tables 3, 4, and 5. Multivariate findings presented throughout the bulletin also draw on these data, as well as other available data for the four trainings being evaluated.

Resources and Barriers

As demonstrated in the current and previous evaluations[2] the balance of organizational resources and barriers is typically a strong predictor of post-training results such as action plan progress. To better assess the impact of organizational resources and barriers, the list of items in the center section of Table 3 was included in the 90 day follow-up. Participants were directed to:

> *Please indicate the extent to which you found each of the following to be either barriers (i.e., lacking, inadequate or problematic areas) OR resources in your organization that hindered OR helped you in applying what you learned in the training to your job.*

Respondents were asked to rate each item in the resource/ barrier list on a seven point scale from –3 to 3 where negative numbers represent barriers and positive numbers represent resources:

–3 = substantial barrier
–2 = moderate barrier
–1 = slight barrier
0 = neither resource nor barrier
1 = slight resource
2 = moderate resource
3 = substantial resource

This scale allowed respondents to distinguish the magnitude of each resource/barrier item. (Note however that no attempt was made to establish the importance of the various items relative to each other, i.e., the resource/ barrier items were not weighted.) Results are summarized in Table 3. Items at the top of the table were more frequently reported as resources, while items toward the bottom were more frequently cited as barriers. For example, participants overall found management/ administrative support to be a moderate resource (1.71). Likewise, participants overall rated staff development/ training a slight resource (1.06). Participants on average rated funding/infrastructure and workload/time the lowest, and thus least likely to be resources and most likely to be barriers.

It is important to note, however, that when standard deviations are relatively high (in this case greater than about 1.0) mean scores near zero do ***not*** indicate that most participants rated the item zero (neither a resource nor a barrier). Instead, this indicates that although the scores averaged out near zero, there was a lot of variation among scores such that the participants who rated the item as a resource approximately equaled those who rated it as a barrier. For example, in the case of workload/time, the mean rating was very near to zero (0.11), and yet only 22% of respondents rated this item as "neither resource nor barrier" (Figure 7). Nearly 40% of respondents found excessive workload or inadequate time to be a barrier to achieving their goals; about 38% reported that adequate time and/or manageable workload amounted to a resource for them. A similar pattern is observed in the distribution of ratings on the funding/infrastructure item (Figure 8).

Post-training Contact with Trainers

To gauge the extent and utility of post-training contact between participants and trainers, one item on the follow-up asked participants to "Please describe the extent to

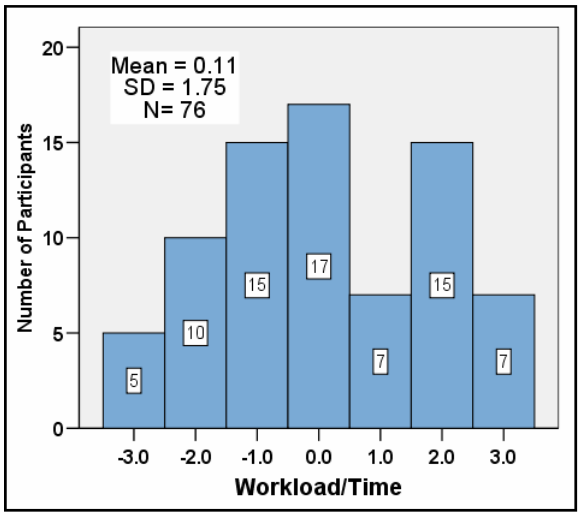

Figure 7

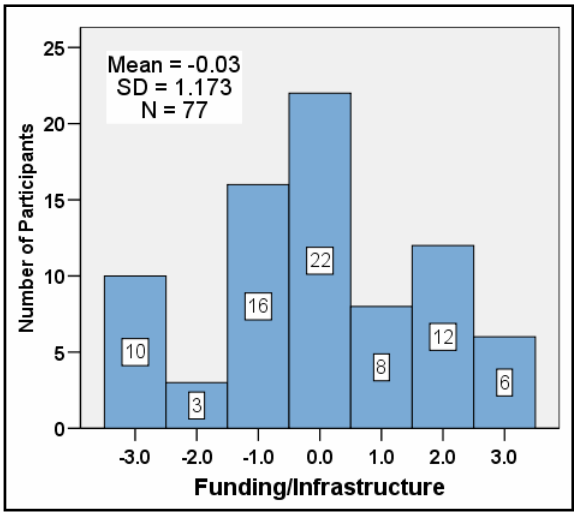

Figure 8

which your trainers or coaches were useful (or not useful) after the training and how so." Content analysis of participant responses indicate that 65.7% of respondents (46 of 70) had some post-training contact with trainers (Table 4). Trainers answered questions, gave advice, or provided documents, materials, or other resources to 20.0% of respondents. Although 47.1% of respondents provided a variety of other positive comments about the trainers, these either directly or indirectly referred to contact ***during*** training, rather than ***after*** training. Of the70 respondents, none provided negative or critical remarks about trainers in response to this survey item.

Lack of Progress on Action Plan Goals

A wide variety of univariate, bivariate, and multivariate findings from the 24 training evaluations reported in the bulletin series have consistently implicated post-training factors, such as organizational resources and barriers, as the best predictors of post-training progress. Post-training and organizational measures typically explain more variation in action plan progress than training measures, such as quality and relevance, and far more than pre-training measures such as personal and demographic variables. Nonetheless, in order to further explore this

Table 4: Post-training Contact with Trainers

	IBM (N=17)		**ASJ** (N=20)		**CPSA** (N=16)		**DSHU** (N=17)		**Overall** (N=70)	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Had post-training contact with trainer(s)	9	**52.9**	18	**90.0**	10	**62.5**	9	**52.9**	46	**65.7**
Trainer(s) answered questions, provided advice, documents, materials, etc.	4	**23.5**	5	**25.0**	4	**25.0**	1	**5.9**	14	**20.0**
Misc. positive remarks about trainer(s)	6	**35.3**	15	**75.0**	6	**37.5**	6	**35.3**	33	**47.1**
Misc. negative or critical remarks about trainer(s)	0	**0.0**	0	**0.0**	0	**0.0**	0	**0.0**	0	**0.0**

Table 5: Lack of Progress on Action Plan Goals **(Bulletin 4, Table 14 Update)**

Content summary of responses to the three-part follow-up item: *For any Action Plan goal in which you did not make as much progress as anticipated: a) What could NIC have done differently in the training to assist you? b) What could you have done differently (during or after the training) to improve your progress? c) What could your agency or organization have done differently to better support you in accomplishing the goal(s)?*

	IBM		**ASJ**		**CPSA**		**DSHU**		**Overall**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
What NIC could have done differently:										
Positive/complementary comments	6	**35.3**	11	**52.4**	6	**40.0**	9	**45.0**	32	**43.8**
Neutral Comments, e.g., "nothing"	13	**76.5**	12	**57.1**	6	**40.0**	16	**80.0**	47	**64.4**
Criticisms and/or suggestions for improvements	1	**5.9**	2	**9.5**	4	**26.7**	3	**15.0**	10	**13.7**
What you (participant) could have done differently:										
Positive/complementary comments	3	**17.6**	3	**14.3**	2	**13.3**	3	**15.0**	11	**15.1**
Neutral Comments, e.g., "nothing"	9	**52.9**	3	**14.3**	2	**13.3**	9	**45.0**	23	**31.5**
Criticisms and/or suggestions for improvements	7	**41.2**	11	**52.4**	7	**46.7**	4	**20.0**	29	**39.7**
What participant's agency could have done differently:										
Positive/complementary comments	1	**5.9**	5	**23.8**	5	**33.3**	4	**20.0**	15	**20.5**
Neutral Comments, e.g., "nothing"	7	**41.2**	2	**9.5**	5	**33.3**	6	**30.0**	20	**27.4**
Criticisms and/or suggestions for improvements	11	**64.7**	17	**81.0**	12	**80.0**	15	**75.0**	55	**75.3**

Responses to the three sub-items above were analyzed together because many participants provided mixed responses rather than strictly limiting each comment to the appropriate sub-item. Percentages may not total 100 because some comments referenced multiple areas. Overall, 17 of 21 IBM participants, 21 of 22 ASJ participants, 16 of 18 CPSA participants, and 20 of 24 DSHU participants provided responses to one or more of the sub-items.

topic, on the training action plan follow-up participants were asked directly:

> *For any Action Plan goal in which you did not make as much progress as anticipated: a) What could NIC have done differently in the training to assist you? b) What could you have done differently (during or after the training) to improve your progress? c) What could your agency or organization have done differently to better support you in accomplishing the goal(s)?*

Results from a content analysis of participant responses are summarized in Table 5. Very few participants (13.7%) reported that NIC could have done anything differently in the training to increase their progress on action plan goals. On the contrary, nearly half (43.8%) of the respondents wrote positive or complementary remarks about NIC. About two-thirds (64.4%) of respondents wrote neutral comments, e.g., "There was nothing that NIC could do to move progress forward." Note that some partici-pants provided multiple comments and/or types of comments, i.e., percentages may not total 100.

Nearly 40% of respondents indicated that they themselves could have done something differently to improve their progress on action plan goals where they fell short of anticipated progress (Table 5). Although responses varied, the most common self-criticism, reported by 14 of the 30 respondents who offered self-criticisms, was that the participant could have been more focused on the goal(s) and/or prioritized better. Only 15.1% of respondents wrote positive or complementary comments about themselves; 31.5% wrote neutral comments.

When asked what their agency or organization could have done differently to better support their action plan progress, 75.3% of respondents provided one or more criticisms or suggestions for improvement (Table 5). Only 20.5% wrote positive or complementary comments, while 27.4% wrote neutral comments. The most commonly cited criticisms or suggestions for better support from their agencies involved workload/time, funding, staff or organizational resistance, and personnel.

Clearly participants overall felt the trainings were of high quality and that there was little more NIC could have done in the trainings to improve post-training action plan progress. This is consistent with the high training satisfaction, learning, and applicability ratings discussed throughout the bulletin series. On the other hand, many participants indicated that they themselves, and especially their agencies, could have taken steps to help improve their progress on action plan goals. The most commonly cited agency criticisms were largely consistent with those previously identified in evaluation findings and discussed throughout the bulletin series, i.e., excessive workload, insufficient funding, and staff resistance. These and other organizational barriers, previously shown to be predictive of post-training action plan progress,[2] may interact with participants' lack of focus or other limitations in non-trivial ways. For example, it is reasonable to suppose that a participant's lack of focus or prioritization may undermine their ability to identify or productively utilize resources, or to overcome organizational barriers necessary to apply their training and achieve their action plan goals.

Table 6: Summary of Findings

Outcomes		Predictors [1]		
		Pre-Training and Demographic Measures	**Training Quality and Relevance Measures**	**Post-Training and Organizational Measures**
Learning (at conclusion of training)	4.55 [a] (SD=.44, N=84)	0.0%	16.1%	
Learning (pre-post change)	1.49 [b] (SD=.78, N=84)	0.0%	28.4%	
Application of Learning (post-training levels)	8.15 [c] (SD=2.10, N=60)	0.0%	27.3%	23.2%
Application of Learning (relative to estimates)	-.77 (SD=1.78, N=48)	55.5%	0.0%	0.0%
Action Plan Progress (post-training levels)	3.05 [d] (SD=1.00, N=76)	0.0%	0.0%	28.6%
Action Plan Progress (relative to estimates)	-.56 (SD=1.07, N=67)	0.0%	0.0%	24.1%

[1] Percentage of variation in each outcome that was explained or predicted by each model (set of predictors); adjusted R^2.

[a] 4.55 represents a high to very high level of self-reported learning at the conclusion of training.

[b] 1.49 units of learning gain during training, where the scale is 1=very low; 2=low; 3=medium; 4=high; 5=very high.

[c] The scale is 0=Not at all; 10=A great deal.

[d] 3.05 represents approximately moderate progress.

Summary and Recommendations

This bulletin was provided as a supplement to Bulletin 4 (*2008 Evaluation Results: Satisfaction, Learning, and Action Plan Progress*). Findings from Bulletin 4 were updated to include recently available follow-up data from CPSA and DSHU. Bulletin 4 findings were expanded to include multivariate results from a series of multiple regression analyses. As a supplement, rather than a replacement, this bulletin is best viewed together with Bulletin 4. The previous bulletin provides necessary background, and is referenced numerous times in the current bulletin.

Data from over 700 completed evaluation forms were analyzed for the bulletin (Table 1). Initial data were provided by 84 training participants and 11 trainers (99% response). Follow-up data were provided by 77 participants (90.6% response). The following FY08 trainings were evaluated:

- Inmate Behavior Management (I08-J2301)
- Administering the Small Jail (08-J2801)
- Conducting Prison Security Audits (08-P3202)
- How to Run a Direct Supervision Housing Unit: Training for Trainers (08-J2202)

Three primary outcomes were examined:

- ***Learning*****,** both post-training levels and pre-post change scores
- ***Application of Learning*** to the job within 90 days of training, and the accuracy of application estimates
- ***Action Plan Progress*** within 90 days of training, and the accuracy of progress estimates

Several hundred variables were considered as potential predictors of these outcomes; approximately 45 were selected, arranged into three models, and tested:

- ***Pre-training measures*** (participant demographics)
- ***Training measures*** (quality, relevance, type, trainer effectiveness, participant attention, etc)
- ***Post-training measures*** (organizational resources and barriers, post-training contact with trainers, post-training application of learning, etc.)

Outcomes are summarized in Table 6, along with the predictive value of each model for each outcome. ***On average, outcomes were quite favorable.*** Learning and post-training application of learning were reported at high levels; action plan progress was moderate. ***However, over-estimation was widespread***; 50% of participants fell short of learning application estimates, and 75% fell short of action plan progress estimates. This pattern may be cause for concern because, for example, estimates often form the basis for commitments to or coordination with colleagues, superiors, other departments, or outside organizations. When actual progress falls short of estimates, reputation may suffer, and relationships can become strained.

Although results for the four "non-estimate" outcomes were quite favorable on average, individual participants varied considerably in their outcomes (Figures 1b-6b). The extent to which each regression model accounted for variation in each outcome is summarized on the right half of Table 6. Several patterns are apparent.

First, the pre-training model was not a significant predictor for five of the six outcomes. This is a very favorable finding in that it provides ***no evidence of bias or discrimination*** in either the training environment, or participants' organizations, based on age, race, gender, or other demographic variables included in the model. The one significant pre-training model appears logical and no cause for concern (see discussion on page 8).

Second, the ***training model*** was more often a significant predictor of ***early or intermediate outcomes,*** while the ***post-training model*** was more often a significant predictor of ***distal or ultimate outcomes.*** Organizational resources and barriers were especially important in the post-training application of learning and action plan progress. These patterns are consistent with previous multivariate findings discussed in Bulletin 3, [2, 32] and further supported by qualitative results from the current evaluation. Content analyses of narrative responses to open-ended items indicate that ***less than 14% of participants believe NIC could have done anything differently to improve their progress on action plan goals.*** About 40% indicated that they themselves could have been more focused or prioritized better. ***Over 75%, however, indicated that their agency or organization could have better facilitated their progress*** in a variety of ways; most comments referenced excessive workload, insufficient funding, and organizational resistance.

Recommendations:

1. Based on the observed pattern of overestimation among participants, future trainings may need to address the issue of conducting realistic appraisals of self and organization in the action planning process.
2. Based on growing evidence of the importance of organizational resources and barriers in achieving favorable outcomes, consider addressing these issues in the trainings, and/or shifting some resources from training, to post-training support.
3. Future research and evaluation should provide for a more thorough and rigorous examination of post-training and organizational factors to identify viable options for improving post-training outcomes.

Future Directions

This bulletin updated, expanded, and concluded findings from the evaluation of four 2008 trainings which we began in our fourth bulletin *(National Institute of Corrections Training Evaluation Project 2008 Evaluation Results: Satisfaction, Learning, and Action Plan Progress).* These bulletins provide very positive evidence that trainer satisfaction with participants, and participant satisfaction, learning, post-training application of learning, and action plan progress were all quite high. On the other hand, these bulletins also reveal some areas where improvements can be made. These areas include assisting trainees in developing a better understanding of training relevance to their organization, as well as helping trainees develop realistic appraisals of their own capacity and the capacity of their organizations to assist with actually implementing and supporting the action plans developed at the training.

The next bulletin, sixth in the series, will present findings from evaluations conducted with established and traditional NIC training courses. It will present training evaluation results from Correctional Leadership Development and Management Development for the Future trainings conducted in the pilot phase of the evaluation project (2005-2006). This bulletin will be unique in that it will present evidence, based in part on data collected from participants' managers, laterals, and direct reports, as to whether or not the trainings had any measurable impact on the trainee's organization.

The final (seventh) bulletin in this series will summarize, synthesize, and extend what we have learned from the training evaluation project to date. Given the amount of data collected over the past few years, and findings they've yielded thus far, we believe we have sufficient evidence to present several theoretical models that will aid NIC developing training support protocols that will make training more effective. To illustrate, when trainees encounter an excess of organizational barriers (or a dearth of resources) that interferes with their ability to apply what they have learned from their training, one theoretical model suggests that new training support protocols could assist trainees in adjusting perceptions of efficacy and adjusting their priorities, thus further enhancing the applicability of the training. Another theoretical model we are developing will attempt to demonstrate relationships between satisfaction, learning, application, and organizational impact, and how NIC, training participants, and their organizations can benefit from this knowledge.

The next series of bulletins will present findings from evaluations conducted with online trainings that NIC is conducting or assisting with. Given the exponential growth of online training in the academic, business and government sectors, we felt it is appropriate to discuss in this bulletin's *Future Directions* why it is important to evaluate online training.

To begin, a definition of online learning is in order. Online learning is the delivery of course content via the Internet. Sophisticated versions of it include video, audio, chat rooms, bulletin boards, frequent assessments, and ongoing documentation. Less sophisticated versions of it are limited to text, slides, written lectures and documents, and web pages.[33]

Both within and outside the field of education, the amount of training conducted online is increasing rapidly. Academic and business environments in the global economy view the implementation of online education and training programs as a necessary avenue for training and implementing programs.[34] Online learning presents many advantages and opportunities for staff development. These include: increased access, greater flexibility, cost savings, and increased collaboration. Despite these advantages, there are several potential disadvantages. They include: quality of content and process, hidden costs, and readiness of the online learner.[33]

There is a lack of conclusive research concerning the effectiveness of online education.[34] This is due, in turn, to a lack of rigorous research methods being used. In general, field experiments of most educational practices are typically weak because they are conducted in settings where it is difficult to control for rival explanations.[35] Research on distance educational practices is no different, since the prevailing view is that it is of low quality.[35] Most research on teaching courses online has primarily been descriptive and exploratory. More recent online research has been unable to provide true experimental data to identify causal relationships.[36] "Continued research is needed to inform learner outcomes, learner characteristics, course environment, and institutional factors related to delivery system variables in order to test learning theories and teaching models inherent in course design."[36]

For these reasons, NIC has decided to play a crucial role in an experiment designed to compare the effectiveness of online versus face-to-face training.

The evaluation will be a between-subjects experimental longitudinal design (see Exhibit 1). The experimental group will involve professional development training courses where online technology is in place. The control group will consist of traditional face-to-face professional development training where online training is not being utilized. Staff will be randomly assigned to the experimental and control groups. Data measures will be collected prior to the training (pre-test), immediately after the training (post-test 1), and several months after the training (post-test 2).

Full experimental design is rarely an option in real world situations. Random assignment, which is often impractical outside of controlled laboratory settings, is the best method to ensure that participants in both groups are equivalent in

terms of knowledge, skill, experience, attitude, etc. In the design of research, the more extraneous differences between the treatment and control groups can be minimized, the stronger the conclusions that can be made.[35] NIC and its partners are providing a rare opportunity to collect hard data about the relative efficacy of online versus face-to-face training.

Sources of data for this evaluation will come from brief, but comprehensive pre-post participant data measures that have been used previously in NIC training evaluations as well as distance learning evaluations. In addition, trainer assessment of participants will also be collected.

One of the primary strengths of this evaluation is that the research design will be built into the development and administration of the training, rather than vice-versa. In this manner, it will be easier to maintain the balance of the instructional impact of the training between the online and face-to-face mediums. Very few, if any, online evaluations have been conducted in conditions as ideal as those we hope to have.

For more information please contact:

Dr. James B. Wells
President and Chief Research Consultant
Commonwealth Research Consulting
4160 Kentucky River Parkway
Lexington, KY 40515

jbwells@cwrc.us
(859) 806-5748

Exhibit 1: Proposed Experimental Design

	Pre-Test	Online Training vs Traditional Training	Post-Test 1	Post-Test 2
Experimental	O	X_a	O	O
Control	O	X_b	O	O

Where:

O = Participant Data Measures

X_a = Training where Online Courses Implemented

X_b = Training where Online Courses Not Implemented

Notes

[1] Some survey forms were combined, i.e., participants evaluated multiple trainers on a single form; trainers evaluated multiple participants on a single form.

[2] Wells, J., Minor, K., and Parson J. (2008, February). *NIC Research Bulletin 3: Training Results, Activity Level Changes, and Implementation Results*.

[3] The location of the bulletins is subject to change. If a search of the NIC website does not locate the bulletins, please contact Dr. James Wells at jbwells@cwrc.us or (859) 806-5748 for copies.

[4] Dillman, D. A. (2000). *Mail and telephone surveys: The total design method* (2nd ed.). New York: Wiley.

[5] Wells, J., Minor, K., and Parson J. (2008, November). *NIC Research Bulletin 4: 2008 Evaluation Results: Satisfaction, Learning, and Action Plan Progress.*

[6] In some cases several versions of each model were tested; the number of independent variables included in each version varied slightly.

[7] See bulletin 4 for additional details on trainer ratings of participants.

[8] Percentage of outcome variation accounted for or explained by the model refers to adjusted R^2. All multiple regression modeling reported in the bulletin was conducted with the enter method.

[9] Adjusted R square, F scores, degrees of freedom, and significance levels are provided in endnotes as appropriate. The full output of multiple regression modeling, including the preparatory and follow-up procedures typically used in support of it, are voluminous and beyond the scope of this publication.

[10] Participant learning was also rated by each trainer; see Bulletin 4, Table 10, and pages 16 and 19.

[11] Adjusted R^2 = .161; $F_{10,55}$ = 2.247, p = .028

[12] Adjusted R^2 = .014; $F_{7,30}$ = 1.077, p = .402 (model not significant)

[13] The leftmost bar in Figure 2b represents all gains of less than 0.5, not only gains of zero.

[14] Adjusted R^2 = .284; $F_{10,55}$ = 3.574, p = .001

[15] Adjusted R^2 = .068; $F_{7,30}$ = 1.384, p = .248 (model not significant)

[16] See Ilian, H. (2004). Levels of levels: Making Kirkpatrick fit the facts and the facts fit Kirkpatrick. In B. Johnson, V. Flores, & M. Henderson (Eds.), Proceedings of the 6th Annual Human Services Training Evaluation Symposium 89-104. Berkeley, CA: California Social Work Education Center.

[17] Adjusted R^2 = .273; $F_{12,38}$ = 2.567, p = .013

[18] Adjusted R^2 = .232; $F_{12,41}$ = 2.332, p = .022

[19] Adjusted R^2 = .019; $F_{7,15}$ = 1.061, p = .433 (model not significant)

[20] t(47)=3.01, p=.004. The mean estimated application at the time of training was 9.05 (N=63, SD=1.35); the mean follow-up application reported 90 days after training was 8.15 (N=60, SD=2.10). However, note that only 48 participants provided both initial estimates and follow-up reports. For those 48 participants the mean difference was –0.77 (SD=1.78).

[21] Adjusted R^2 = .021; $F_{12,28}$ = 1.072, p = .418 (model not significant)

[22] Adjusted R^2 = -.020; $F_{5,35}$ = .844, p = .528 (model not significant)

[23] Adjusted R^2 = .555; $F_{7,10}$ = 4.033, p = .023

[24] Adjusted R^2 = .286; $F_{12,59}$ = 3.372, p = .001

[25] Adjusted R^2 = .134; $F_{15,43}$ = 1.599, p = .114 (model not significant)

[26] Adjusted R^2 = .039; $F_{7,23}$ = 1.173, p = .356 (model not significant)

[27] t(66)=4.31, p<.001. The mean estimated progress at the time of training was 3.58 (N=67, SD=0.73); the mean follow-up progress reported 90 days after training was 3.02 (N=67, SD=1.03). The mean difference was –0.56 (SD=1.07).

[28] Adjusted R^2 = .163; $F_{15,35}$ = 1.650, p = .110 (model not significant)

[29] Adjusted R^2 = -.039; $F_{7,17}$ = .871, p = .549 (model not significant)

[30] Clearly the application of training-related learning to ones job is not purely "effort" in the strictest sense of the word; application can reasonably be interpreted to include an element of "results". For example, efforts to apply learning may be rebuffed by organizational resistance, in which case the participant may report a lower application rating. Similarly, action plan progress is not purely "results" in the strictest sense of the word; it also includes an element of "effort". Nonetheless, for illustrative and comparative purposes it is reasonable to characterize learning application as primarily "effort" and action plan progress as primarily "results".

[31] Adjusted R^2 = .241; $F_{7,30}$ = 2.677, p = .028

[32] The smaller percentages of explained variance in the current evaluation, relative to the evaluation described in Bulletin 3, stems at least in part from the smaller datasets available for the current evaluation, i.e., approximately 20-60 cases, depending on the variable, compared to 100-300 cases for the evaluation described in Bulletin 3. Nonetheless, it is the similarity in the pattern of findings that is important.

[33] Killion, J. (2000, October). Online staff development: Promise or peril? *NASSP Bulletin, 84*(618), 38-46.

[34] Bartley, S.J., & Golek, J.H. (2004). Evaluating the cost effectiveness of online and face-to-face instruction. *Educational Technology & Society, 7*(4), 167-175.

[35] Bernard, R.M., Abrami, P.C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P.A., Fiset, M., & Huang, B. (2004, Fall). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, 74(3), 379-439.

[36] Tallent-Runnels, M.K., Thomas, J.A., Lan, W.Y., Cooper, S., Ahern, T.C., Shaw, S.M, & Liu, X. (2006, Spring). Teaching courses online: A review of the research. *Review of Educational Research, 76*(1), 93-135.

This page intentionally blank.