

SECTION TWO: ASSESSMENT, DIAGNOSIS AND EVALUATION

Five guiding principles to help improve diversity training assessment

Abstract

Purpose – Provides clear guidelines to diversity training practitioners to help improve assessment of training. Encourages crosstalk between academics and practitioners.

Design/methodology/approach – Reviews some of the research on the benefits versus costs of diversity training assessment and generates five core principles to help practitioners identify and exploit assessment opportunities.

Findings – Most diversity training initiatives are neither routinely nor systematically assessed, in spite of there being clear business benefits from doing so, such as improved diversity management, enhanced organizational efficacy, and increased responsiveness to diversity needs. Suggests reasons for the lack of assessment, such as lack of an obvious payoff for business, suspicion and fear of what assessment might reveal, and lack of experience among practitioners of how to optimally assess their initiatives. Provides five core principles to guide practitioners through the process of assessment: deriving testable hypotheses; obtaining baseline data or using naturally occurring control groups to get an index of change; ensuring assessment measures appropriately tap goals of training and training, itself; considering short and longer term assessment approaches and taking into account the wider organizational context.

Practical implications – Enables diversity training practitioners to engage with the process of assessment, a topic that receives very little attention in spite of the widespread use of diversity training as a means of enhancing diversity management.

Social Implications – Discusses an important problem: the lack of systematic appraisal of diversity training. Better assessment techniques will lead to more accurate knowledge about diversity training outcomes which will, in the long run, enhance diversity management.

Originality/value – Bridges the gap between the academic work on this topic and practitioners' needs for clearly articulated ideas to help them put theory and research about assessment into practice.

Article type – Conceptual/review paper

Keyword(s) – Diversity training; Assessment; Evaluation

Introduction

Diversity training (DT) is currently the main diversity management tool in organizations. A recent survey of individuals within UK organizations charged with diversity management responsibilities found that 94% of respondents said their organization employed DT awareness training, and 77% mentioned offering manager DT (CIPD, 2006). With our combined experience as both academics and DT practitioners, we have previously reported how little mainstream social science and management theory and research finds its way into the development of DT programs (Pendry, Driscoll, & Field, 2007). In the present article, we turn our attention to the topic of putting DT *assessment* into practice. In simple terms, when and how can practitioners incorporate what we know from the social science and management literatures to better assess the effects of DT?

In discussing why assessment is not a more routine part of diversity management, we provide a clear picture of hurdles that can deter a practitioner from including assessment in their work. We hope that acknowledging some of these hurdles - and discussing how to circumvent those hurdles that can't be jumped serves as an impetus to practitioners to try assessment.

In this paper we:

1. Show how assessment can be achieved more wisely and directly by genuinely helping businesses develop meaningful benchmarks and metrics.
2. Provide organizational illustrations to help make our points more tangible and applicable to both practitioners and organisations interested in managing diversity.
3. Demonstrate how DT practitioners might better assess the impact of DT interventions through a deeper appreciation of empirical research and theory.

What the benefits and costs of diversity assessment?

Thousands of organizations offer diversity training, yet few seem to routinely evaluate and disseminate their assessment findings. The result is that we have little idea if these programs are successful. When diversity training has been systematically addressed, though, real benefits obtain. We outline a few of the more consistently obtained benefits.

Benefit 1: Assessment can increase organizational efficacy

Organizations that assess initiatives will be more effective in what they do (Hubbard, 2003). This is partly because assessment determines whether real progress is being made. Without knowing outcomes, it is difficult to benefit from what is working (or improve it), as well as difficult to fix, modify, or stop what is not working.

Benefit 2: Assessment can highlight where organisations need to take corrective action

As assessment clarifies *what* needs to be modified, it allows organizations to be responsive. For example, if we find that only white males in the organization are negative about developing and retaining diversity and it correlates with their negative feelings about their own advancement and promotion, then corrective steps can be taken to educate via training (if there is a lot of misinformation), promote transparency about current recruitment and advancement practices (if there is a need for clarity), and/or even review/investigate recruitment and advancement procedures (if there is reason to suspect abuse).

Benefit 3: Assessment helps garner support for diversity

Having an accurate picture of the benefits/costs of diversity can help management and employees understand the rationale for the direction the company is taking. Materials gathered for assessment purposes, such as the organization's demographic profile across different job levels or survey items that document harassment levels among employees, can show why diversity initiatives are needed.

Benefit 4: Diversity assessment helps create more diversity competency organizations

As assessment improves responsiveness and garners support for diversity practices, the benefits of pursuing and managing diversity are achieved by the organization, such as improved recruitment and retention of employees, better performing work teams, a warmer work climate and reduced litigation (Cox & Blake, 1991; see Table 1).

We are mindful, too, of the costs of assessing diversity training initiatives that may understandably dissuade practitioners from routinely engaging in it and organisations from requiring it. Our experiences suggest most of these costs can be minimized or avoided.

Cost 1: It is an added financial cost and organisations may not consider that sufficient benefits will accrue from budgeting for DT assessment

Although organisations often embrace diversity management, and see some benefits of training (e.g., reduced liability, improved retention of women/ethnic minorities), they may not wish to spend money on assessment initiatives when it is neither mandatory nor immediately obvious why it is advisable. Worse, what happens when training is not showing an immediate effect? Or backfires? Although many practitioners are motivated by a concern to do DT well, for some, the decision to omit outcome measurement may be "...born out of a fear of knowing" (Hubbard, 1997, p. 12). The norm is that organizations are not held accountable for demonstrating specific benefits of diversity training, only for engaging in training. In order to validate an assessment approach, practitioners need to make assessment normative and demand more accountability from any diversity practitioner they hire

Cost 2: Diversity assessment can elicit suspicion both for practitioners and employees

Some diversity practitioners are suspicious of diversity assessments because they feel the diversity field is being held to a higher standard than other fields and the object of doing the assessment is to justify program/budgetary cuts. Likewise, some employees are suspicious of diversity assessment. They feel coerced to answer in a way that is supportive of existing work practices because of fears for job security. Such suspicions could be allayed if it was made clear that individual data was not released but was being handled by an outside consultant/firm, and further, that one's individual responses are not identifiable to the organization. We suspect there would be more support for diversity training and assessment, and fewer well-intentioned but disastrous actions, if suspicions were met head-on or waylaid in the first place with more transparency about the organization's diversity goals, methods, and assessment process.

Cost 3: DT practitioners may not be experienced in assessment

Unless practitioners have a background in assessment, they may not be able to evaluate the legitimacy of assessment instruments. They may even try to create assessment instruments, themselves, leading to poor measures (i.e., surveys with leading questions, double barreled questions, etc.) and then make inaccurate conclusions (i.e., statements of causality). Some practitioners may not even be aware of the assessment literature and potential benefits of assessment. Those who are assigned diversity as part of their HR job specification, but have no background in DT, may feel out of their depth. A recent CIPD survey found that 53% of respondents charged with responsibility for diversity management activities do not consider themselves diversity specialists (CIPD, 2006). For such individuals, assessment may not be at the top of their 'to do' list. Better understanding of, and training in, evaluation methods would be beneficial.

Given all of these costs, it is unsurprising that DT assessment rarely happens. Although we are sensitive to these potential costs, we would argue that training that allows for meaningful assessment leads to organizational benefits, as well as ultimately improves the accountability of the diversity training field. In the current economic climate, this becomes even more important.

Guiding principles: Learning from existing research and practice

Our extensive review of the literature (available upon request) leads us to propose five principles that will help DT practitioners learn from existing research and practice. We outline these in the following section, and where appropriate, offer a workplace illustration of how these principles may be put into practice.

1. What do you hope DT to DO? Derive testable hypotheses and assess

Ideally, we need to know not just *if* DT can work, but *how*, and *why*. Accordingly, a specific hypothesis (an assumption about what training will achieve) in DT research will not merely assert that the DT programme will "work", but rather that, under particular

organizational conditions (e.g., support from upper management, clear communication of diversity as a value, a company that is experiencing growth), certain manipulations (specific *methods* of intervention) will have certain consequences (specific, measurable *outcomes*).

Putting this principle into practice

To illustrate, consider an organization that is concerned about higher turnover rates for women and ethnic minorities and would like to improve retention rates for these groups (i.e., current versus desired *organizational condition*).

The agreed training *method* is awareness building for managers, where knowledge about barriers to women/ethnic minorities that lead to their exit from organizations, and methods of overcoming such barriers, are highlighted, explained and encouraged to be put into practice

Here, the DT *hypothesis* would be:

Managers attending training will better understand why women/ethnic minorities leave an organization and will be better at implementing retention programs that lead to the improved retention of female/ethnic minority employees in comparison to managers who have not yet undergone DT.

Outcomes of training that demonstrate support (or not) for this hypothesis can be assessed by employing measures that:

1. Tap, in the short term, into trained (vs. not) managers' clearer knowledge of barriers and successful retention efforts put into place since the training and
2. Show, in the long term, more retention strategies put into place by trained (vs. not) managers and any associated successes with decreasing turnover of women and under-represented ethnic minorities.

This approach to DT is rarely the case for several reasons:

1) *Most DT does not develop in this quasi-experimental way.* It is rare that such an explicit, upfront hypothesis is found within the DT literature and indeed, in practice. It is fairly common for authors to include some description of certain organizational conditions but much rarer to find a detailed, explicit description of training methods, specific effects expected from training, or how these three elements might relate to each other. Trainers, similarly, often do not approach DT in this way, emphasising instead the delivery of a product that is fit for a given purpose (without paying heed to organizational conditions, hypotheses, methods of DT and incorporating tangible outcomes into their design).

2) *The norm is to provide training, not to assess it.* Some organizations are simply following the prescriptive norm: to provide DT that is minimally disruptive, in line with what other organizations are doing and what is required by law. Such a norm simply calls for training to be offered, not for proof it is working. As noted earlier, this norm needs to change if assessment is to become a more routine part of the DT package.

3) *Assessment measures do not tap goals of DT.* When organizations do have specific DT goals, can develop testable hypotheses, and do try to incorporate assessment, the outcomes of training are not always assessed appropriately. Instead, the focus of the assessment is upon trainees' immediate reactions to the training/trainer. Practitioners should consider how best to marry together aims, hypotheses, methods and outcomes of DT.

4) *Trainers may not know what has prompted the decision to provide training.* Trainers are sometimes brought in without knowledge of previous DT initiatives or specific incidents prompting training (e.g., a discriminatory email joke resulting in a lawsuit). Sometimes, even upper HR administrators within the organisation are unaware. Such contextual information is vital to avoid compromising the ability of the trainer to link goals of training, hypotheses, manipulations, and likely outcomes. DT practitioners are well-advised to push organizations for such information if not given a chance to do their own

5) As soon as possible, a practitioner should view the organization's most recent cultural audit, or conduct one, prior to planning DT. This acquaints practitioners with knowledge about the *organizational conditions* noted earlier. Although not always possible, it is good practice to remind the organization that without knowing the issues and understanding ongoing diversity activities at the organization, neither training – or assessment of the training – will be optimal.

Where the training and evaluation materials are mutually agreed upon between organization and practitioner, it may be possible to develop specific, testable hypotheses. This process is summarised in Figure 1.

Insert Figure 1 here

Take home point: Training that evolves by looking at the current organizational climate, pinpointing where change is desirable, developing clear hypotheses about what training might achieve, and incorporating some means of assessing this, will, ultimately, be more informative for practitioners and more impactful for organizations.

2. How do you measure if change has occurred? Try to obtain baseline data and/or use naturally occurring control (no training) groups

If one wishes to assess change following DT, one has to ask: Change in relation to what? There are two methods to tackle this. First, one can obtain pre-test data on certain issues, and then repeat these measures post-training, to see if change has occurred. This repeated

measures method has been used in published evaluation studies within organizations (e.g., Hanover & Cellar, 1998). If this is not possible, one can ask trainees on two separate scales – one marked “Before the training...” and the second marked “After the training...” to “List” or “Rate” your understanding of...why there may be a cold climate for women; why retention rates are lower” etc. Although this method can be compromised by demand effects, it allows for some comparison pre/post training in those cases where it has proved impossible to gather baseline data.

A second method is to incorporate control group procedures, to be sure that any changes observed in a trained group are not also observed in participants who do not undergo the training (e.g., as a result of wider cultural change within the organization). However, such quasi-experimental approaches are not the norm. There are clearly logistical and ethical reasons for this in applied settings. For example, where DT is mandatory, legal issues may make it impossible to refrain from delivering training to all employees. Even where DT is not mandatory, companies do not want to be held liable (i.e., having to explain in court why a manager accused of racial harassment did not receive training because s/he was in the control group).

However, the fact remains that it is challenging to make claims about DT’s effectiveness when one is not confident that outcomes are solely due to the training. One suggestion is to exploit naturally occurring control groups whenever possible.

Putting this principle into practice: Naturally occurring groups

To give an illustration, an organization seeks wholesale training of its workforce, but financial or practical constraints necessitate that DT is only provided to a representative sub-section of the workforce, initially, and rolled out to the rest of the workforce later.

In such cases, one can compare a host of measures for those sub-sections of the workforce trained vs. not.

Where practicable one should aim for random assignment to treatment/control groups since a failure to do so can compromise findings. For example, confining initial training to only managers and then rolling training out to the rest of the workforce creates problems of comparison between trained managers (the treatment group) vs. not trained non-managers (control group). Using a control group that is a consequence of rolling out training over time is not a perfect method; nonetheless, it can still provide important information about DT outcomes.

Take home point: Our experiences in the field make us only too aware that most training does not incorporate repeated measures methods or control groups because it is just not possible to do so or organizations do not wish to. However, practitioners and organizations keen to understand more about the effects of DT would be well-advised to consider ways to exploit opportunities that effectively capture change as a consequence of DT.

3. How do you assess what you have done? Aim for consistency across goals, actual training, and outcome measures

What happens when goals, training and assessment do not match up: Many organizations do not match training needs to the training undertaken, or either one of these to assessment measures. This can lead to problems of assessment, as the following example demonstrates:

Putting this principle into practice

To provide a concrete illustration of what can occur when there is a mismatch between goals, training and assessment, a university may seek to create a ‘warm’ environment for all students, regardless of gender, race, etc

A training goal should therefore be to show the teaching staff how to create a warmer classroom environment. Instead, staff are taught about how ethnic groups tend to be different from one another in the classroom (i.e., women are more cooperative; men are more competitive, etc.)

Classroom climate is then assessed

Here, the training goal and assessment measures match, but the actual training had little to do with the goal. Not surprisingly, the impact of training upon climate is negligible.

A mismatch can occur for many reasons. Sometimes, the organization and practitioner do not communicate. For example, practitioners may not be cognizant of the problem of not matching goals to training to assessment, especially if they were not privy to the specific goal that was formulated or outcome(s) desired. Such goals or desired outcome(s) may have only been articulated within the university administration and never communicated to the practitioner. Another reason for the mismatch occurring is that practitioners may have little understanding of the attitude literature, and instead act upon their naive hypotheses about how just raising awareness about groups should lead to changes in behavior.

Is training intended to foster increased awareness, changes in diversity attitudes, or behavioural intentions?: Moving on to the measures themselves, there are many to choose from. One’s choice of measure(s) should derive from goals and content of the training itself. Measures vary and may include:

- Affective scales (e.g., How do you feel about the notion of workplace inequality?)
- self-reported behavioural intentions (e.g., When you meet a new colleague from a minority group in the future, will you change anything about the way you interact with them?)
- changes in actual skills/behaviour (assessed either by observation or later self-report)

- knowledge/awareness (e.g., What do you think the 2010 Equality Act means to organizations?)
- perceived importance (e.g., How important are diversity management practices in this organization?)

Remember your ABCs: There are several ways practitioners might distinguish between different possible DT outcomes, and methods of assessment. One common method is to focus upon the ABCs: **A**ffect (feelings towards) **B**ehaviours (actions towards), and **C**ognitions (for example, thoughts, knowledge, and beliefs about). Such a distinction is found within the social psychological literature on attitude structure (e.g., Eagly & Chaiken, 1993) and also finds its way into management training evaluations more generally (Kirkpatrick & Kirkpatrick, 2006) and DT evaluations more specifically (e.g., Hayles, 1996). Many existing evaluation studies employ one or more of these measures as having measures that corroborate gives more confidence in findings.

Allied to these points, social psychologists have argued that practitioners should pay closer attention to intervening mechanisms (e.g., the psychological processes that underlie observed prejudicial behaviour) when considering the kinds of bias they wish to address/assess in DT (Dovidio, Gaertner, Stewart, Esses, & ten Vergert, 2004). For example, where interventions focus upon how participants *feel*, they typically have more success in reducing bias in behaviours with a strong feeling (affective) component (such as how positive you feel toward or your willingness to engage in contact with members of other groups). In contrast, where interventions focus more on participants' *thoughts*, this may have greater impact upon associated cognitive processes (e.g., views on public policy). In sum, a better understanding of how affect, behaviour, and cognitions impact, alone or in concert, upon bias can make for more effective DT design and assessment.

Beware of a mismatch between goals and assessment measures: This point fits within the broader recommendation that the choice of measure(s) ought to be driven by the goals of DT and the actual training. Some of the published evaluation studies we have seen are not systematic in this respect. For example, Tansik and Driskill (1977) sought to change attitudes via DT, and employed role play/empathy building techniques (i.e., active behaviours) to promote change. However, their intervention method may not have mapped so well onto the outcome measures (responses to ethnic labels/semantic differentials that tapped attitudes towards different groups; example – To what extent do you think Chinese people are...tick a point on a seven point scale anchored 'shy' and 'outgoing'). Such a measure is likely to have elicited long-held beliefs about groups rather than any positive affect or behaviour elicited by the role playing/empathy building techniques. A more direct outcome measure could have looked at whether there was an increase in positive and/or empathic behaviour toward the groups. Practitioners need to recognize the importance of such matches in content/outcome.

Draw appropriate conclusions: More generally, researchers sometimes draw conclusions about DT outcomes without having critical details about the types of interventions employed. For example, in their survey, Naff and Kellough (2003) focused upon outcome measures at the organizational level, comparing ratios of promotion, dismissal and

resignation between groups since they viewed such measures as acceptable proxies for employment equity. They did not scrutinise the type of DT that organizations had offered (e.g., quality, quantity, methods) instead simply asking organizations to say whether or not they offered training. So, one organization might tick the box for offering training when it was confined to a one hour seminar/video presentation whereas another might be offering a several day workshop with background readings/active discussions. For the purposes of analysis, both programs would be deemed equivalent. Simply noting training presence/absence, though, could obscure pertinent findings. Correlating these data with ratios of promotion, dismissal and so on does not permit us to deduce with confidence how types of training impact upon these distal factors. Why change may have occurred and its sustainability is not clear and presents a challenge for practitioners trying to make inferences from the literature to the applied setting and interventions at their disposal.

Learn what scales are out there and choose carefully: It is beyond the scope of the current article to review in detail the huge array of assessment measures in use. Many that are available are often contextual and or/focused upon knowledge/attitudinal/skills aspect. For example:

- Multicultural Awareness Knowledge Skills Survey (MAKSS; D'Andrea, Daniels, & Heck, 1991) and the Multicultural Counselling Inventory (MCI; Sadowksy, Taff, Gutkin, & Wise, 1994) are designed with clinical settings in mind.
- The Instructor Cultural Competence Questionnaire (Roberson et al., 2002) asks trainees to respond to a series of hypothetical diversity incidents (vignettes that might require modifying to suit different organizational settings).
- Within the social psychological literature, too, there are specific scales that tap attitudes towards certain groups (e.g., Modern Racism Scale: McConahay, 1986; Modern Sexism Scale: Swim, Aiken, Hall, & Hunter, 1995).

Before using any of these scales, practitioners must be clear how well they fit the hypotheses, goals and training in question and adapt as appropriate or create their own measures.

Be creative – go beyond scales if permissible: For certain training objectives, for example those that are not to do with discrete, manifest constructs, one might take unobtrusive observational measures and assessments of diverse workgroup productivity and creativity. Similarly, if training aims to improve interaction with diverse groups, one might combine measures of behavioural intent, third party-rated behaviour, employee self-reports of cultural competence and so forth. Trainee self-reports in isolation may sometimes be inconsistent with their actual behaviour and may be prey to social desirability effects. Hence, converging methods of measurement may tap more aspects of a construct and increase confidence in the appraisal. It should be noted that such

interventive methods are not always practicable or popular with employees, so trainers should proceed with caution if considering them.

Take home point: In sum, there are many measures to consider using in training. Deciding which are appropriate will be determined by paying close attention to the hypotheses, goals and types of training undertaken. The crucial point is, where possible, to try to consider assessment as an integral part of the overall package of DT design and delivery.

4. When do you need to assess DT: Short and longer term considerations

Beware the problems of relying upon immediate assessment: Another concern with assessment is the *times* at which DT effects are measured. As noted, often outcome measures are taken only once, immediately after training. There are some problems with this practice. First, it means that self-report measurements are prey to demand characteristics (i.e., if distributed by and given back to the trainer that one is evaluating). Irrespective of the manner in which the data is collected, participants presented with self-report items that refer to specific subject areas within the training may infer what the practitioner is looking for, and what constitutes a socially desirable (“good” or “bad”) response. Positive results are in this sense unsurprising – and not wholly persuasive with regards to deeper changes. It is a starting point, though, and appropriate for shorter, introductory trainings where a practitioner may simply want to establish that certain information presented was understood and there are no consensual problems. Second, by taking measures only once, just after training, there are no baseline measurements available for comparison. Thus, for any assessment of change, it is desirable (though not always practicable) that baseline measurements are also taken.

Consider if longer term assessment is feasible: Where possible, follow-up assessments of DT should occur over time to establish the exact nature of any changes. Obviously, practical demands may drive the feasibility of this option.

Putting this principle (almost) into practice

Hanover and Cellar’s (1998) study tried to capture longer term responses to DT for certain measures. Ultimately, however, organizational policies obliged them to take these additional measures rather sooner than they felt was optimal. Indeed, the measures were taken only two months post-training, a timescale decided upon in order to allow the control group rapid access to training.

This illustrates the obvious tensions that we appreciate may exist between a desire for longer-term assessment of certain issues, and the constraints of organizational demands.

Nonetheless, it is desirable, where practical, to assess possible changes in outcomes obtained over a suitable time period (given the expected impact of the training).

For example, it may be the case that changes are transient or delayed, both trends we would not capture with a single post-training measure. For interventions that run on over several weeks, more frequent assessment across the intervention period may be desirable to tap changes. Equally, not all outcomes suit immediate testing. It may take a long time before newly acquired diversity skills can be harnessed and demonstrated. In DT, it is often the case that complex information and suggestions about attitudinal/behavioural change are advanced with the hope that in later real-life intergroup interactions, people will be able to remember and employ skills introduced. This is a big 'ask' and will require time and the opportunity to practice. As such, detecting behavioural changes in the short term is unrealistic whereas a longer term assessment may bear fruit if action plans to practice one's learning are implemented as well.

Look to the future: Looking further ahead still, if one wishes to tap longer term organizational outcomes of DT, such as comparing ratios of promotion, dismissal and resignation, one should do so after a meaningful period of post-training time has elapsed (and as noted above, to do so with close reference to the types of DT offered so that outcomes can be closely matched to training content and goals).

5. What about the bigger picture? Take account of other factors that might affect DT impact

Thus far, we have considered how DT might be better assessed, and in doing so, have honed in on the micro-level aspects of programme evaluation. In addition to the assessment points already made, it is sensible to step back and consider how other factors may impact upon DT outcomes.

Consider the organization's diversity management ethos: This may impact upon the efficacy of training, and where possible, it makes sense to gather data on this (e.g., via a cultural audit). If the organizational climate is not broadly supportive of diversity management, and DT is being conducted simply to tick an Equal Opportunities box, then this may impair DT efforts. It will also be detrimental to the organization if assessment takes place, then no training, or training with no follow up (action plan, change in policies, etc.). Raising expectations for change and then not implementing any changes is, at best, demoralizing and will discourage or anger employees from under-represented groups. The best designed programme will inevitably fail if employees work within an organization that only pays lip service to diversity management issues. The social psychological literature on intergroup contact cautions that prejudice reduction is most likely where there are *social norms of equality*. In the workplace, this equates to an organizational climate that creates and reinforces a norm of acceptance and tolerance (see: Pettigrew & Tropp, 2000). If this is not the case, practitioners can use such information to, for example, build the case for diversity (during training sessions and meetings) and emphasize what needs to be in place to improve the organization's diversity competency.

Think about the demographics and differing motivations of DT trainees: It is important to assess how these may impact upon DT outcomes. Demographic data can be useful,

especially if a particular demographic category is underrepresented in (or within particular levels in) the organization. Collecting potentially sensitive demographic information in a non-reactive way is a vital part of understanding workplace diversity dynamics. Preferably, such demographic information is already on file and available to permit an investigation of survey responses by demographic group. The right to use employee demographics for such a purpose, though, varies, with laws about its use differing from country to country. Having to collect demographic information from individuals can lead to resistance/reactance (even changes in their self-report of their demographics, like putting down a different racial group) and raises suspicions of how their demographics will be used which can affect their survey responses (Driscoll, Zawojewski, & Stahura, 2008). Practitioners have to be responsive to such reactions.

Putting this principle into practice

Roberson et al.'s (2009) evaluation study illustrates why such demographic data and knowledge about organizational characteristics can enrich DT assessment procedures. They examined certain individual and environmental factors that might influence the use of skills transfer strategies following DT. They took measures of:

Ethnicity data (these were collected via the graduate training office of the university in which the study took place).

Organizational characteristics which were assessed using Burke and Baldwin's (1999) four item scale which taps one's immediate supervisor's responses to skills taught, together with situational cues (data on proportion of people of colour at upper levels of the organization).

By gathering data on trainee ethnicity, and organizational characteristics, it was possible to gauge how DT efforts are helped or hindered by such variables.

They found that trainee race/ethnicity was a significant predictor of the extent to which skills were transferred following DT. More specifically, Non-White trainees were more likely to try to use their training back on the job, in comparison to White trainees. Management support and organizational culture were also found to significantly affect transfer strategies.

Overall these findings underscore how important it is to consider DT efforts within the broader organizational context and with reference to the demographic make up of the DT attendees.

Develop situational measures further if appropriate: Depending upon the organization, such data might, for example, comprise a blend of self-report items (where trainees are asked to provide views on what they perceive the ethos to be, perhaps provide anecdotal examples) together with more objective data on organizational policies and procedures regarding factors such as career progression, child-friendly working hours, maternity/paternity leave arrangements, bullying, harassment, and so forth. Such data

can identify potential weaknesses or inconsistencies in either the existence or delivery of certain relevant policies that might impede DT impact. For example, one might find a general and elaborate paternity leave program, but anecdotal evidence showing men are afraid to use paternity leave and low rates of usage. The pattern of findings across various measures then suggests the policy is not functioning as intended.

Mind the gap: Care should also be taken to note any training-unrelated experiences that may impact upon employees' responses in the time that intervenes between training and follow-up assessments. This risk is inevitable when working within an active business, but recording such events can help explain unexpected results. An additional measure in post-training assessment probing employees' training-unrelated experiences, or even an open-ended question on a survey "Is there anything important that you'd like to add?" can give insight and is one we recommend. Ideally a range of 'other' factors should be controlled or at least recognized for their independent role in either assisting or hampering DT initiatives. For example, if there are job searches that are filled by employees from under-represented groups – whether coincidentally or not -- during the intervening time period, such hires may impact upon post- training assessments.

Take home point: Do not neglect to take account of other factors (related to the organization's ethos, employees' characteristics, or other non-training related experiences that may affect assessment) and be prepared to react to these issues and modify training to ensure it is fit for purpose.

Points for thought

At this point we wish to recap. What are some key points for thought that arise from the present analysis? Essentially, there are several ways in which practitioners might improve the assessment of DT:

- We suggest practitioners press organizations to clearly specify their DT goals (i.e., what they hope DT will achieve) and work with the organization to derive testable hypotheses.
- At the same time, practitioners might develop creative ways of assessing if the organizations' goals have been met in the short and longer term.
- Rather than sole reliance upon reactionnaires, practitioners could collect baseline data and return at an agreed upon time to collect longer-term reactions.
- Practitioners might also remain alert to the possibility of using naturally occurring control groups if they can't randomly assign to experimental and control groups.
- Practitioners ought to conduct any assessment with a view to matching hypotheses, training, and outcomes to maximize training effects, as well as using multiple assessment measures to corroborate any given finding.
- Finally, practitioners should be familiar with the prevailing ethos of the organizations they are working with, previous cultural audits and DT trainings, employees' current needs, and how to better control for the potential desirable or undesirable effects of existing policies and practices upon DT outcomes.

References

Burke, L. A., & Baldwin, T. T. (1999). Workforce training transfer: A study of the effect of relapse prevention training and transfer climate. *Human Resource Management, 38*, 227-242.

CIPD (2006). *Diversity in business: How much progress have employers made? First findings*. London: CIPD.

CIPD (2010). Diversity: An overview. Retrieved from <http://www.cipd.co.uk/subjects/dvsequel/general/divover.htm>

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *Academy of Management Executive, 5*(3), 45-57.

D'Andrea, M., Daniels, J., & Heck, R. (1991). Evaluating the impact of multicultural counselling training. *Journal of Counseling and Development, 70*, 143-150.

Dovidio, J. F., Gaertner, S. L., Stewart, T. L., Esses, V. M., & ten Vergert, M. (2004). From intervention to outcomes: Processes in the reduction of bias. In W. G. Stephan & P. Vogt (Eds.), *Education Programs for Improving Intergroup Relations: Theory, Research, and Practice* (pp. 243-265). New York: Teachers College Press.

Driscoll, D. M., Zawojewski, J. S., & Stahura, J. (2008). Learning from engineering students' reactions to small group laboratories. In Zawojewski, J. S., Diefes-Dux, H., & Bowman, K. (Eds.) *Models and modelling in Engineering Education: Designing experiences for all students* (pp. 111-136). Rotterdam, the Netherlands: Sense Publishers.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. New York: Harcourt, Brace, Jovanovich.

Hayles, V. R. (1996). Diversity training and development. In R. L. Craig (Ed), *The ASTD training and development handbook: A guide to human resource development*, 3rd ed., (pp. 104-123). New York: McGraw-Hill.

Hubbard, E. E. (1997). *Measuring diversity results*. Petaluna CA: Global Insights Publishing:

Hubbard, E. E. (2003). *The Diversity Scorecard: Evaluating the Impact of Diversity on Organizational Performance (Improving Human Performance)*. Oxford, UK: Elsevier Inc.

Intercultural Development Inventory (IDI): Applications with Individuals, Teams, and Entire Organizations. Retrieved 10 May 2009 from http://www.mdbgroupinc.com/idi_applications.htm

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels*, 3rd ed. San Francisco, CA: Berrett-Koehler. 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.

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.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-126). New York: Academic Press.

Naff, K. C. & Kellough, J. E. (2003). Ensuring employment equity: Are federal diversity programs making a difference? *International Journal of Public Administration*, 26(12), 1307-1337.

Pettigrew, T. & Tropp, L. (2000). Does intergroup contact reduce prejudice? Recent meta-analytic findings. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp.93-114). Mahwah, NJ: Lawrence Erlbaum Associates.

Roberson, L., Kulik, C. T. & Pepper, M. B. (2001). Designing effective diversity training: influence of group composition and trainee experience. *Journal of Organizational Behavior*, 22, 871-885

Roberson, L., Kulik, C. T., & Pepper, M. B. (2002). Assessing instructor cultural competence in the classroom: An instrument and a development process. *Journal of Management Education*, 26, 40-55.

Roberson, L., Kulik, C. T., & Pepper, M. B. (2009). Individual and environmental factors influencing the use of transfer strategies after diversity training. *Group Organization Management*, 34, 67-89.

Rynes, S. & Rosen, B. (1995). A field survey of factors affecting the adoption and perceived success of diversity training. *Personnel Psychology*, 48, 247-270.

Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the multicultural counseling inventory: A self-report measure of cultural competencies. *Journal of Counseling Psychology*, 41(2), 137-148.

Swim, J. K., Aikin, K. J., Hall, W. S., Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68, 199-214

Tansik, D. A. & Driskill, J. D. (1977). Temporal persistence of attitudes induced through required training. *Group & Organization Studies*, 2(3), 310-323.

Table 1: Benefits of Pursuing Diversity

Dimension of Interest	Possible Organizational Benefits Pursuin
Recruitment of Talent	Easier to recruit employees
Retention of Talent	Easier to retain employees
Teaming	Better performing teams (innovative, more divergent thinking, etc.)
Work Climate	Warmer work climate (more cooperative, respectful, etc.)
Lawsuits, Litigation, Grievances, & Complaints	Decrease in number and severity (more effective responding to discriminatory situations, better remedies, greater fairness in disciplinary measures, layoffs, transfers, etc.)
Markets & Customers	New markets open; new customers attracted
Productivity/Profitability	Increased productivity/profitability a by-product of other benefits

Figure 1: Diversity training assessment as a natural process arising from dialogue between practitioner and organization

