

2012

Training Evaluation in Virtual Worlds: Development of a Model

Richard N. Landers

Old Dominion University, rlanders@odu.edu

Rachel C. Callan

Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_fac_pubs

 Part of the [Community Psychology Commons](#), and the [Industrial and Organizational Psychology Commons](#)

Repository Citation

Landers, Richard N. and Callan, Rachel C., "Training Evaluation in Virtual Worlds: Development of a Model" (2012). *Psychology Faculty Publications*. 16.

https://digitalcommons.odu.edu/psychology_fac_pubs/16

Original Publication Citation

Landers, R. N., & Callan, R. C. (2012). Training evaluation in virtual worlds: Development of a model. *Journal For Virtual Worlds Research*, 5(3), 1-20. doi:10.4101/jvwr.v5i3.6335



Volume 5, Number 3
Managerial and Commercial Applications
December 2012

Training Evaluation in Virtual Worlds: Development of a Model

Richard N. Landers

Rachel C. Callan

Department of Psychology
Old Dominion University, VA, USA

Abstract

Many organizations have adopted virtual worlds (VWs) as a setting for training programs; however, research on appropriate evaluation of training in this new setting is incomplete. In this article, we address this gap by first exploring the unique issues relevant to evaluation faced by training designers working in VWs. At the macro-organizational level, the primary issue faced is an organizational culture unreceptive to or otherwise skeptical of VWs. At the micro-organizational level, two major issues are identified: individual trainees unreceptive to VWs and general lack of experience navigating VWs. All three of these challenges and their interrelationships may lead to poor reactions, learning, and transfer from VW-based training despite strong, pedagogically sound training design. Second, we survey the training evaluation research literature, identifying the most well-supported training evaluation models, discussing the suitability of each for evaluating VW-based training. Third, we propose a new integrative model based upon this literature, incorporating solutions to the unique issues faced in VWs with the most relevant portions of the models discussed earlier. Fourth, broad thematic implications of this model are identified and applied to prior VW literature. Finally, we provide specific recommendations to practitioners and researchers to evaluate their VW-based training fully.

1. Introduction

In the late 2000s, training managers began to identify virtual worlds (VWs) as a promising setting for training and development. At IBM, the VW Second Life was used as part of new employee training and in developmental activities (Hatch, 2007). Sun Microsystems recreated its United States physical campus in Second Life, using it for training and more (Wyld, 2010). In *Harvard Business Review*, Reeves, Malone and O'Driscoll (2008) recommended VWs as more effective in teaching soft leadership skills than traditional leadership training alone.

The embracing of VWs as a training setting by early adopters can be credited to two major advantages (Wyld, 2008). First, the tools provided in most VWs enable a greater degree of interactivity and illustration than is easily accomplished in traditional training settings. Participants can direct their avatars to participate in role plays with greater realism than in a traditional workshop, and other complex activities and ideas can be simulated with great detail. For example, a real-life oil rig might be recreated in a VW so that trainees can explore it at zero risk to themselves or the functioning of the rig (Gronstedt, 2007). Second, costs are much reduced. Participants do not need to travel, time away from the work desk is reduced, and the virtual facilities themselves are relatively inexpensive in comparison to traditional training facilities (Wyld, 2010). If desired training outcomes are equally strong between traditional training and training in VWs, VWs emerge as the superior choice.

Despite these apparent advantages and the surge in press coverage of a handful of high profile training programs, VWs have not yet overwhelmingly replaced traditional training settings. This may be because current research has provided relatively little guidance on effective instructional design in such environments (Lee & Kim, 2010). However, the American Society for Training and Development's (ASTD, 2011) annual State of the Industry Report probes why so-called "Web 3.0" technologies, including artificial intelligence and VWs, face slow adoption. At the top of the list, 76.4% of the 412 surveyed organizations identified "budget constraints" to be a barrier. In second place, 64.3% identified "lack of understanding from leadership".

Superior learning outcomes and reduced costs in comparison to traditional training are perceived by researchers to be the primary advantages of training in VWs. These advantages, if evaluated appropriately, should make a compelling argument to corporate leadership. Unfortunately, they are often not communicated effectively. Most current research on learning in VWs takes place in educational settings, and this research literature does not comprehensively examine outcomes that would be valuable to organizations or consider organizational characteristics that would affect these outcomes. Although several models of training evaluation are available, these are not commonly used in the current research literature involving VWs.

Given this gap in the literature, in the present article, we explore how to more comprehensively evaluate training taking place in VWs. In doing so, we identify how well-designed training in a VW may still appear ineffective in an organizational context if it is not evaluated appropriately, ultimately leading to under-adoption. To fill this gap, we first explore the unique challenges to evaluation involving VWs. Second, we explore the major current models of training evaluation with the goal of identifying the particular value of each model to the VW context. Third, we combine the most critical portions of each of these models into a parsimonious integrative model. Finally, we explore the implications of this new model for the previous research literature on VWs.

2. Unique Challenges Faced When Evaluating Training with a Virtual World Setting

The challenges faced when creating training programs in a VW are not very different from those faced when creating training programs anywhere else. The learning activities chosen must be pedagogically sound, based upon well-validated learning theory, and professionally delivered (Mamo et al., 2011). In this way, VWs are only another wrench in the toolkit of the training designer, to be used when effective realization of learning objectives suggest them (e.g., when simulation is the preferred instructional technique) or budget restrictions necessitate them.

However, designers face different challenges when attempting to evaluate the results of such training programs. Many employees and their managers are skeptical of participating in VWs, which is likely to negatively affect training outcomes. In a survey of 81 training professionals, Tayler and Chyung (2008) found that the majority of respondents (60%) were either undecided or against the use of VWs in training. One respondent commented, “The 3D training environments do not fit into our organizational culture” (p. 23) and another commented, “I just don’t see the value.” (p. 23). After a three year study of business uses of VWs, Bessière, Ellis and Kellogg (2009) concluded that “the hardest challenge we faced in carrying out our studies was convincing employees to participate” (p. 2886), with employees “viewing it as a distraction at best, and potentially risky or immoral at worst” (p. 2886). Additionally, VWs are considered “merely games” by many business executives (Gartner, 2007), echoing findings that students often consider VWs to be play spaces rather than a valuable setting for learning (Cheal, 2009), despite increasing evidence that games and gamified processes can be valuable for learning (Landers & Callan, 2011). Furthermore, recent meta-analytic research has identified a positive correlation between utility reactions and learning outcomes ($r = .17$; Blume, Ford, Baldwin, & Huang, 2010), supporting an association between poor attitudes toward the value of VWs for training and reduced learning. Altogether, this prior research suggests some resistance to and discomfort with VWs from all organizational members, from managers to employees, and additionally that this discomfort is associated with negative outcomes. This resistance might be driven by the individuals within the organization or by the larger organizational culture to which they belong, problems that would each require a dramatically different approach to solve.

Additionally, even those employees who are comfortable enough to experiment with VWs often find the experience frustrating due to lack of experience or training in VW navigation. Bessière, Ellis and Kellogg (2009) found that employees had a difficult time navigating and interacting with the VW as they intended, on top of a wide variety of technical difficulties, echoing the challenges faced in educational uses of VWs (Bedford et al., 2006; Shen & Eder, 2009). Most users assumed that the VW software would be intuitive to use, skipping reading training sheets on VW navigation provided to them. This assumption proved unwise; VW navigation is itself a complex skill.

The critical issue faced is that these issues – skepticism of VWs and difficulty learning the software among trainees – do not necessarily reflect poorly upon VW-based training or VW in general. For example, consider a training designer implementing a VW-based training program. After putting in a great deal of time and effort developing and deploying the VW and training program, she gets her first satisfaction surveys back afterward only to discover that her trainees hated every minute of it! With an appropriate evaluative model, she would soon discovered the real cause of this apparent failure, enabling her to redirect her energy toward correcting the real problem: poor pre-training attitudes toward VWs, an organizational culture that does not support the use of VWs, or a lack of trainee experience in 3D virtual environments. However, we suspect that in most cases, the designer is more likely to interpret poor post-training satisfaction surveys as either 1) an indicator of a failed training design or 2) a reason not to use

VWs in future training efforts. These conclusions may not be justified. An identical training program, completed with trainees experienced with and accepting of VWs, might have been a rousing success.

In summary, an evaluative model is needed that considers training outcomes in a practical, straightforward manner, but also incorporates pre-training attitudes and experience. Thus we explore the dominant training evaluation models in the research literature with an eye toward these issues.

3. A Review of Training Evaluation Models

Most central to the concept of training evaluation is that emotional/affective reactions to training are only one component of comprehensive evaluation. These survey-based satisfaction measures are the most common form of training evaluation by far, but do not correlate strongly with learning or the transfer of learned skills to the job (Brown, 2005). Thus, the goal of formal training evaluation models is to more comprehensively describe training outcomes desirable to organizations.

Broadly, there are three types of training evaluation models: process models, hierarchical models, and mediational models. Process models focus on the training designer's role in evaluation, describing important elements of the decision-making process. For example, a process model might link organizational goals pre-training with evaluative goals post-training. Hierarchical models structure the outcomes as a series of interrelated measurement targets within trainees. For example, such a model might posit that training outcomes of interest can be described as knowledge, skill, or affect. Mediational models are the most complex of the three, proposing a series of causal relationship between trainee characteristics and organizational objectives, often incorporating elements of both process models and hierarchical models. We will next describe dominant models from each category in turn, highlighting their particular relevance (or lack of relevance) to training taking place in VWs.

3.1 Process Models of Training Evaluation

Alvarez, Salas and Garofano (2004) developed a broad process model of training evaluation and effectiveness, beginning before training has been designed, and represent this category well. In this model, needs analysis (a formal information gathering stage that precedes training design) directly affects three major processes: the design and content of the training program, changes in the learners, and organizational results. Individual differences between trainees are hypothesized to influence each of these processes, whereas the design of the training is hypothesized to affect changes in learners and organizational results, and organizational characteristics are hypothesized to affect organizational results. This model is considered comprehensive because it includes all relationships consistently found in a review of 73 training articles over 10 years.

Although this model is an excellent basis by which to understand the macro-organizational level issues related to training within organizations, there is little we would expect to be different for training designed to take place in a VW. A needs analysis should still precede adoption of a training program, and individual differences (generally) should still affect how well that needs analysis results in various broad outcomes. The researchers suggested that this model supersedes many classic models because it is so broad. The lack of specificity in this model makes it broadly useful across a wide variety of training scenarios, but its incremental value in the VW context is limited as a result. Although this model certainly applies to training in the VW context, it does not offer much value in extending VW-specific theory. Other process models are similarly limited; there is little reason to think that the linkages between organizational objectives and training design should typically be different for VW-based training and we will therefore consider them no further.

3.2 Hierarchical Models of Training Evaluation

Perhaps the most influential model of training evaluation is the hierarchical model proposed by Kirkpatrick (1976). The Kirkpatrick model is appealing for its simplicity. It characterizes training evaluation as the assessment of four outcomes at two levels of measurement, making it accessible for practitioners in virtually any context. At the measurement level of the individual, it proposes: Level 1 (Reactions), Level 2 (Learning), and Level 3 (Behavior). At the measurement level of the organization, it proposes Level 4 (Results). Level 1 encompasses affective responses, like those captured by satisfaction surveys. Level 2 includes all cognitive and skill-based outcomes. Level 3 examines resulting behavioral change on the job. Level 4 focuses upon macro-organizational outcomes, like return on investment or effect on stock price. The simplicity of this approach enables easier communication of training effectiveness to those outside of the training field than more complex models involving a greater number of variables can. Thus, this aspect of the Kirkpatrick model is attractive as a solution to the challenge of communicating the value of VWs.

Kirkpatrick (1976) proposed a hierarchy in his model such that higher-level outcomes depend upon lower-level outcomes; that is, 1) to see organizational benefit, trainees must use their newly learned skills on the job, 2) for trainees to use their skills on the job, they must have learned the material during training, and 3) for trainees to learn the material, they must have a positive affective reaction to the training. “Reactions” (Level 1) are the most common outcome examined in corporate settings, which practitioners somewhat disparagingly refer to as “smile sheets”. Kirkpatrick theorized that reactions were critical to other training outcomes. If employees did not enjoy the training, they would be unlikely to maintain attention on the trainer, with negative effects on all higher level outcomes.

The Level 2 outcome in the Kirkpatrick model is “learning.” As practitioners began to adopt this model for evaluating their own training, learning was most often measured with a single multiple-choice paper test, but it is important to note that the model as Kirkpatrick framed it included more than just changes in knowledge. Under this model, learning includes changes in knowledge, skills, as well as attitudes. If a bank manager acquires knowledge about new laws related to his job, an emergency response worker is taught how to perform CPR, or a teacher comes to respect differing viewpoints during diversity training, all of these employees have reached Level 2 in the Kirkpatrick model. Because there are so many kinds of learning, there are just as many types of evaluations one can perform to identify how much learning has taken place. Common evaluative techniques for learning include written post-tests and skill demonstrations administered immediately after training.

Managers and employees learn in training with the goal of transferring gained knowledge, skills, and attitudes to their own work, which describes Level 3 of the Kirkpatrick model, labeled “behavior”. Successful transfer is one of the most difficult goals for a training designer and is itself modeled as several components (Goldstein & Ford, 2002). Transfer is commonly considered the most important goal of training (Baldwin & Ford, 1988) and as a result measuring transfer should be an integral part of training evaluation. Generally, transfer is assessed by evaluating job performance relevant to the training content, but surveys are also sometimes used to ask how often trainees use what was learned on the job.

Finally, in the Kirkpatrick model, behavioral change is necessary for subsequent desirable organization-level changes at Level 4, “results”. Kirkpatrick emphasized that organizations do not solely provide training so that their employees might learn. Organizations provide training in order to reach higher-level organizational goals through that learning. For example, upper management may want to improve managerial job performance related to teamwork in order to improve collaboration within and

output of their work teams. Thus, this ultimate goal of organizational value added is the most prominent distinction between organizational evaluation models and those typically found in education.

The Kirkpatrick model has been heralded for its simple approach to training evaluation. However, training within VWs includes more complex relationships than this model allows; although the process models discussed earlier include room for individual differences like motivation or experience with VWs, the Kirkpatrick model does not. Given this, the simplicity of four outcome dimensions seems valuable in creating a model of VW-based training effectiveness and should be incorporated into a VW-focused model. However, a balance must be struck between simplicity and explanatory power; the Kirkpatrick model is not quite descriptive enough.

Kraiger, Ford and Salas (1993) attempted to improve the precision of the Kirkpatrick model by building upon Kirkpatrick's (1976) split of learning outcomes: cognitive, skill-based, and affective. Cognitive outcomes include not only verbal knowledge but also the organization of this knowledge and cognitive strategies learned in training. The category of skill-based outcomes included not only learning the skill itself but also the process of learning to perform the skill efficiently through the cognitive processes of proceduralization and composition. The last outcome category, affective outcomes, includes both attitudinal and motivational change. The Kraiger model includes self-efficacy and motivational disposition as an outcome while including another aspect of motivation, goal-setting.

Although this model provides a useful hierarchical structure for categorizing outcomes, like the Kirkpatrick model, it lacks individual differences that may be critical to the VW context. Further, Alvarez and colleagues (2004) noted that although Kraiger, Ford and Salas (1993) conceptualize learning outcomes more comprehensively than Kirkpatrick (1976), the model is too impractical to apply to training in the field and has not been noticeably adopted by practitioners as a result. This again emphasizes the need for a practical model of training in VWs. Although we recognize the overall value of the Kraiger model in more accurately describing training-related learning outcomes, we do not see it providing substantial incremental value in the evaluation of VW-based training specifically beyond other models and will therefore consider it no further.

3.3 Mediation Models of Training Evaluation

Mediation models generally represent a blend of both process and hierarchical models. These models tend to be the most balanced between explanatory power and simplicity because they incorporate both practical process-oriented components and detail-oriented hierarchical components. They are also structured temporally and causally, such that the effects of one part of the model can be predicted in later parts of the model. For example, a model might hypothesize that better training design leads to better transfer, but only if learning is also improved. We identified two major mediation models.

Noe and Schmitt (1986) built a theoretical model of training evaluation driven by motivational influences with the inclusion of several other individual differences that were hypothesized to influence these relationships. This model included locus of control which was thought to influence reaction to skill assessment, trainee expectancies, and career/job attitudes. Noe and Schmitt noted previous models did not emphasize trainee expectancies, which they theorized are crucial for motivating employees to learn. This motivation at least in part was hypothesized to determine learning, which in turn influenced behavior and results. Likewise, motivation to transfer was included in the model, moderating the relationship between learning and behavioral outcomes. These additions to the model are important because they suggest that trainees are complex, with a myriad of motives and beliefs that can cause them

to at times be successful in training and others less successful, which speaks well to our goals in developing a VW-focused model.

The final piece of the Noe and Schmitt (1986) model is environmental favorability. This term encompasses both social support for transferring the knowledge, skills, and attitudes learned in training to the workplace but also includes the extent to which employees have the necessary tools and resources to complete the tasks required by their jobs. In the model, this directly influences motivation to learn, motivation to transfer, and training results. Noe and Schmitt recognized that even with well-designed and otherwise effective training, if trainees know their peers or supervisors view the training negatively, the trainee will be less motivated to learn the material or to transfer it, which would diminish organizational-level outcomes (Level 4, in Kirkpatrick's model). Those seeking to train employees in VWs should be aware of the attitudes of all employees if they want to realize result-level outcomes.

Although the Noe and Schmitt (1986) model includes many of the individual differences and environmental factors that may play a role when training in the VW, it is not without its detractors. The biggest issue with this model is it is too complicated to be useful for practitioners training in the VW. The model certainly accounts for most of the important aspects of training evaluation, but to measure each of these variables would be cumbersome, and perhaps unnecessary to provide effective evaluation. Additionally, Noe and Schmitt found mixed support for their complete model, suggesting that it may be more complex than necessary for parsimonious measurement.

Baldwin and Ford (1988) devised a model that addressed many of the limitations associated with Noe and Schmitt's (1986) model and the Kirkpatrick (1976) model. The Baldwin and Ford model focuses on the factors leading to the transfer of training, conceptualizing the process as inputs and outputs in the training program system. It then makes specific hypotheses as to the causal effects of those components. Thus the overall structure of the Baldwin and Ford model is that of a process model with hierarchical elements contained within, making it quite practical. Training designers can see everything that "goes into" training and the anticipated effect of each component.

This model contains three broad categories of training inputs that uniquely affect training outputs: the characteristics of the trainees, the design of the training, and the work environment. Trainee characteristics include ability, personality, and motivation. Training design is comprised of principles of learning, sequencing of the training, and the content of the training. The work environment facet includes the support provided by co-worker and managers but also opportunities to put what is learned in training to use once back on the job. Each of the training inputs is uniquely related to different training outputs. All three inputs are hypothesized to affect learning and retention directly. Trainee characteristics, learning/retention, and work environment then uniquely affect transfer. In this way, training design only affects transfer through its mediating effect on learning/retention.

The empirical evidence for Baldwin and Ford's (1988) model is generally strong. Recent meta-analytic findings confirm that individual differences such as cognitive ability, personality traits, and motivation do in fact broadly predict learning outcomes, as Baldwin and Ford proposed (Colquitt, LePine, & Noe, 2000). Another meta-analysis supports the relationship between individual differences and transfer, as well as work environment and transfer (Blume, Ford, Baldwin, & Huang, 2010). Although environmental factors such as organizational climate do directly affect training outcomes, they also do so indirectly through trainee motivation to learn (Colquitt et al., 2000). This is not represented in the Baldwin and Ford model and should be critical to any model regarding training evaluation in VWs for reasons that we will highlight in the next section.

The Baldwin and Ford model thus combines the appealing simplicity of the Kirkpatrick model (1976) with the individual and environmental factors suggested by Noe and Schmitt (1986) that may be particularly influential when evaluating training in the VW. One portion of the model must be updated to reflect current meta-analytic evidence, and it does not currently contain any VW-specific constructs, but it is otherwise an ideal starting point for a new, integrative model.

3.4 A Comparison of Models, Past and Present

To maximize the theoretical and practical value of training evaluation models in the VW context, an integrative model is needed that comprehensively describes training outcomes and their antecedents, incorporating both micro- and macro-organizational constructs important to VWs while also being both parsimonious and practical. The table below summarizes the strengths and limitations identified in each of the models discussed in relation to these goals, and the discussion below describes the relationship between these strengths and limitations and our goals as stated above.

<i>Model</i>	<i>Summary of Strengths</i>	<i>Summary of Limitations</i>
Alvarez, Salas and Garofano, 2004 (Process)	Comprehensively describes decision-making process for a training designer; considers macro-organizational factors	No unique implications for VW context
Kirkpatrick, 1976 (Hierarchical)	Simple four-part taxonomy of outcomes; easily applied	Does not consider specific individual differences relevant to VW context; narrowly defined
Kraiger, Ford and Salas, 1993 (Hierarchical)	Better consideration of psychological dimensions of training outcomes	Impractical to apply in the field; does not include individual differences relevant to VW context
Noe and Schmitt, 1986 (Mediational)	Explores relationship between trainee characteristics over times before, during, and after training; considers some macro-organizational characteristics	Impractically complex; not parsimonious; empirical support has been mixed across the various components of the model
Baldwin and Ford, 1988 (Mediational)	Combined strengths of process and hierarchical models; straightforward and easily applied	Does not highlight VW-specific concerns; does not include effects of macro-organizational characteristics on individuals

Alvarez, Salas, and Garofano (2004) developed the most comprehensive process model available, and it considers both micro- and macro-organizational factors that influence training success. Because the model is so broad, it can be used to understand training design in nearly any context. However, this characteristic also implies that we would not expect the processes it describes to be related differently in the VW context. For example, as stated in this model, a needs analysis should always drive decisions

regarding training design, regardless of other factors. A model of evaluation within the VW context will add the most practical value by including factors that are uniquely valuable within that context, so this framework is not broadly useful for our purposes here. However, we do note the importance of macro-organizational factors stressed by Alvarez et al. (a feature missing from most other popular evaluation models) and work to include this in our integrative model.

Similar to Alvarez et al. (2004), the narrow nature of Kirkpatrick's (1976) model limits it for extending theory in the VW context. Although it is broadly applicable, the Kirkpatrick four-factor model of training outcomes cannot provide much guidance specific to the VW context. However, its continued popularity among training practitioners highlights the attractiveness of a relatively simple and memorable set of outcomes. Kraiger, Ford, and Salas (1993) expanded the "Level 2: Learning" outcome of the Kirkpatrick model to encompass multiple dimensions, but it has not generally replaced the Kirkpatrick model due to its limited practicality. Thus, from the hierarchical models, we conclude that a simple, straightforward set of outcomes at both micro- and macro-organizational levels will be most valuable for inclusion in an integrative model. Kirkpatrick's model itself should be integrated as the outcome portion of the integrative model to maximize its practicality.

In developing their mediational model, Noe and Schmitt (1986) take into consideration many of the individual differences and organizational characteristics ignored by Kirkpatrick (1976) and Kraiger, Ford, and Salas (1993). Like Alvarez, Salas, and Garofano (2004), Noe and Schmitt (1986) aimed to provide researchers with a comprehensive model of training, including characteristics and outcomes that would apply to nearly any training context. Again, this generalizability results in little value specific to the VW context. Echoing concerns about the Kraiger et al. model, Noe and Schmitt's (1986) model is also too complicated for training designers to utilize or researchers to investigate effectively, a problem compounded by its mixed empirical support. From this, we conclude that even among mediational models, a parsimonious model with a relatively small number of high-importance relationships is needed.

The final mediational model by Baldwin and Ford (1988) combines the strengths of process and hierarchical models into a mediational model that is not as complex as Noe and Schmitt (1986) but is more strongly empirically supported. Thus, the Baldwin and Ford model can be considered the most parsimonious model available to describe the general context of training evaluation. With this in mind, we base our new integrative model primarily upon this model, extending and altering it only insofar as it adds unique value to understanding training evaluation within the VW context (maintaining parsimony) or improves its practicality.

4. Integrating Training Evaluation Models for Training in VWs

Given the balance of simplicity and practical advantages to Baldwin and Ford's (1988) model, we base our new model principally upon it. To be more consistent with recent meta-analytic research on the effect of work environment discussed earlier (Colquitt et al., 2000), we first update the Baldwin and Ford model to include work environment as a cause of intermediary trainee motivation. To maximize the practicality of our model, we replace Baldwin and Ford's (1988) outcome labels with those of the Kirkpatrick (1976) model by reframing Baldwin and Ford's "Learning and Retention" and "Generalization and Maintenance" in terms of that model, changing these to "Reactions and Outcomes" and "Behavioral Change and Organizational Results", corresponding to Levels 1 and 2, and Levels 3 and 4, respectively. Because reactions do not correlate strongly with transfer (Brown, 2005), we distinguish between effects related to Learning and Reactions. Based upon the work of Bessière, Ellis and Kellogg (2009) discussed earlier, we third add two VW-specific constructs: attitudes toward VWs

and experience with VWs. Finally, we hypothesize five new relationships not contained within the original Baldwin and Ford model implied by these changes.

The full model is depicted in Figure 1. Solid black arrows indicate relationships contained within the original Baldwin and Ford model for which there is strong empirical support, whereas dashed lines with numbers represent the new relationships hypothesized here, which we will discuss below.

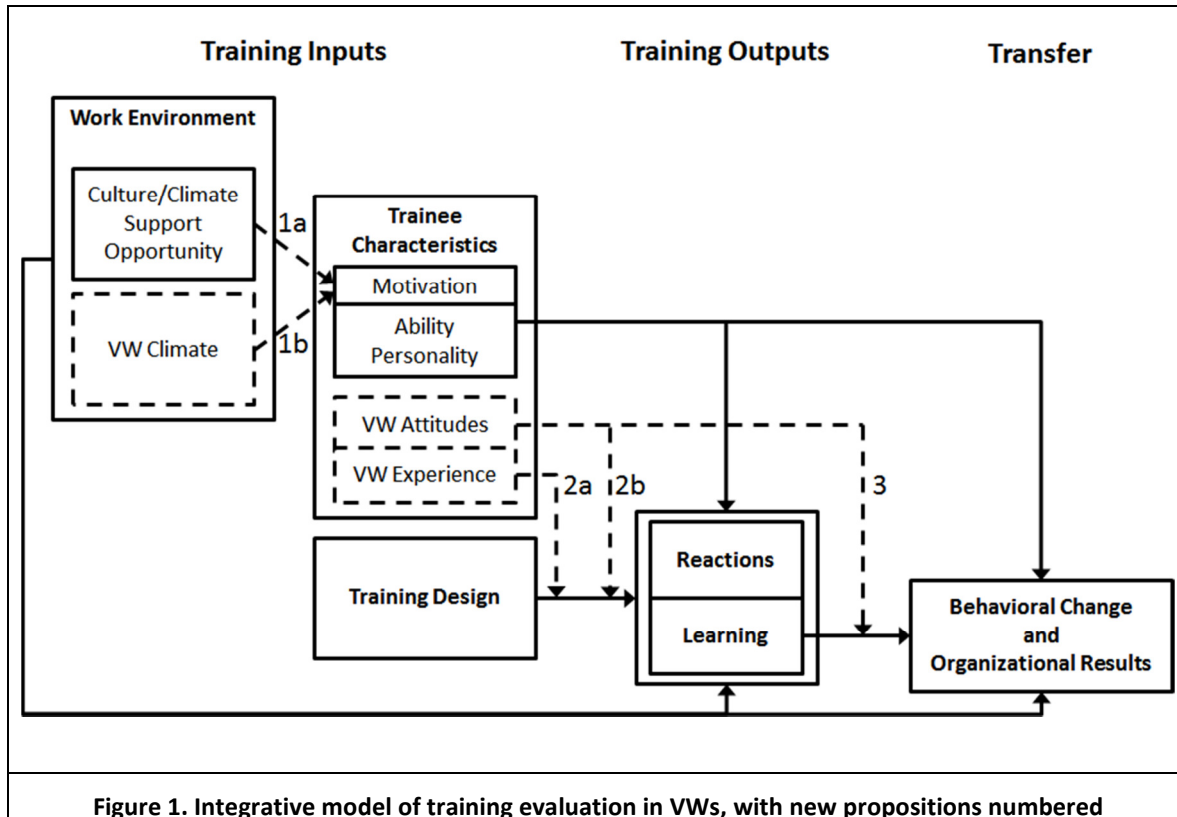


Figure 1. Integrative model of training evaluation in VWs, with new propositions numbered

4.1 Development of Propositions

Proposition 1a. *Work environment characteristics, including organizational culture or climate, supervisor support, and opportunity, affect training outcomes through their mediating effect on motivation to learn.*

Overall, Proposition 1a posits that the work environment affects training outcomes through its effect on trainee motivation. In their meta-analysis of 106 articles within the training literature, Colquitt and colleagues (2000) identified that the relationship between organizational climate and learning or transfer was partially mediated by motivation to learn. In other words, organizational climate was theorized to affect trainee motivation to learn, which in turn affected learning and transfer. Although this cross-level relationship (the effect of work group and organizational attitudes on trainee characteristics) has not previously been included in training evaluation models, Proposition 1a is a well-established finding in the training literature at large.

Proposition 1b. *VW-related work environment characteristics, including VW climate, affect training outcomes through their mediating effect on motivation to learn.*

In Proposition 1b, we apply the reasoning of Proposition 1a to the VW context. This mirrors the structure of Proposition 1a but focuses upon the effect of poor organizational climate toward VWs on

training outcomes. We again base this upon the meta-analytic linkage supported by Colquitt and colleagues (2000) discussed above. Work environment characteristics influence the motivations of employees in a wide variety of contexts, and there is no reason to suspect that VW-related attitudes differ in this.

Additionally, there is some early evidence within the training technology literature that a negative climate toward training technology can influence training outcomes. Landers and Goldberg (*in press*) found that negative supervisor attitudes toward a social network site deployed as a training tool resulted in a very low adoption rate despite initially favorable impressions and intentions among employees. Because supervisors communicated a low perceived value of the social network site to employees (i.e. contributed to a poor climate), even initially enthusiastic employees stopped using it, negating any potential benefit.

Proposition 2a. *Experience with VWs moderates the relationship between training design quality and outcomes. That is, the relationship between training design and training outcomes will be stronger when trainees have more experience with VWs.*

In Proposition 2a, we suggest that the relationship between training design and reactions/learning is moderated by trainee experience with VWs. We root this proposition in cognitive load theory (Sweller, 1988) which is based upon research suggesting that the human brain has limited resources available to dedicate to learning. In the classic metaphor of cognitive psychology, the brain is like a computer: there is a finite amount of processing power available at any given moment, and once those resources are fully engaged, attention must be sacrificed in one area to improve attention in another.

Trainees without experience navigating VWs will devote a portion of their cognitive effort during training toward navigation and exploration, distracting them from training objectives (Mania, Wooldridge, Coxon, & Robinson, 2006). This is consistent with Sweller's (1994) description of the effects of extraneous load, which describes the extra burden placed on the learner by an instructional design that requires directed learner attention. If learners must concentrate on navigating a VW, that attention will not be directed toward the instructional content, resulting in poorer outcomes.

Proposition 2b. *Trainee attitude toward the value of VWs for training moderates the relationship between training design quality and outcomes. That is, the relationship between training design and training outcomes will be stronger when trainees have more positive attitudes toward VWs.*

In Proposition 2b, we suggest that the relationship between training design and reactions/learning is moderated by trainee attitudes toward VWs. When trainee attitudes toward VWs are positive, the quality of training design will have a positive impact on learning. When trainee attitudes toward VWs are negative, the quality of training design will have less or no impact on learning. We root this in both the theory of constructivist learning and adult learning theory, each of which assert that learners must actively create their own knowledge and skills (Bruner, 2009; Knowles, Holton & Swanson, 2005). These are presented in opposition to so-called objectivist learning theories, which rest on the assumption that the goal of training is to deliver an objective body of knowledge to learners as efficiently as possible (Kraiger, 2008).

From this, we conclude that if a learner has negative attitudes toward VWs, that learner is less likely to value the information presented in that context when constructing their own knowledge. As a result, learners with negative attitudes will benefit less from a high-quality training design than will learners with positive attitudes.

Proposition 3. *Trainee attitude toward the value of VWs for training moderates the relationship between learning and transfer. That is, the relationship between learning and transfer will be stronger when trainees have more positive attitudes toward VWs.*

In Proposition 3, we argue that even if learning has occurred, transfer of learning to the workplace from knowledge gained in a VW is less likely to occur if that learner has negative attitudes toward VWs. That is, we propose that no matter how much trainees learn in a VW-based training program, meaningful behavioral and organizational change cannot result if trainees have extremely poor attitudes toward VWs (e.g., “Nothing I learned from a computer game can be valuable in the real world.”). In their review of the literature, Grossman and Salas (2011) noted one of the primary barriers to the transfer of learned skills is the perceived value of training. If trainees do not perceive a link between required performance in a training task and performance-related outcomes that they value, they are less likely to apply any knowledge gained to their jobs (Burke & Hutchins, 2007). Velada, Caetano, Michel, Lyons and Kavanagh (2007) further demonstrated that the match between training requirements and job requirements predicted transfer. We conclude from this that if a trainee has negative attitudes toward VWs and does not believe the skills gained in the VW map onto “real skills” to be used on the job, transfer is unlikely, even if learning has occurred.

5. Applying the Integrative Model

To explore how the proposed model informs the study and practice of training in VWs, we will next investigate how this model might reframe conclusions drawn in prior training studies in VWs. To accomplish this, we considered the proposed model and current research literature, identifying an initial list of common situations where variables highlight by this model could influence conclusions. Through discussion and additional review of the literature, we iteratively refined this list to four major categories of research practice of concern: quasi-experimental comparisons of dissimilar groups, unmeasured organizational and individual antecedents of training performance, limited or missing VW navigational training, and incomplete exploration of the outcome domain. Each of these major themes will be discussed below with examples from the literature and implications of these practices.

It is also important to note that this is not intended as a criticism of prior foundational work in the study of training in VWs; instead, our purpose is only to highlight how the proposed model illuminates observed differences between these studies in an effort to explain their sometimes discrepant or negative results.

5.1 Quasi-Experimental Comparisons of Dissimilar Groups

Quasi-experimentation is a necessary evil in field research. Many researchers publishing in the VW-based training literature incorporate quasi-experimental designs because these are the designs enabled by the convenience samples available to them, such as students enrolled in courses. Although such designs limit the ability of researchers to draw causal conclusions, such conclusions can still be made if threats to internal validity are explained and explored. However, when the researchers fail to measure important individual differences, there is the potential to unknowingly have dissimilar groups in each condition which may confound the results. This is often addressed by the inclusion of control variables or covariates, but prior theoretical and empirical work is needed to identify which control variables are needed (Atinc, Simmering, & Kroll, 2012; Spector & Brannick, 2010). From the proposed model, the individual differences that are most important for use as covariates and controls are personality, ability, and motivation to learn. If work groups or other higher level groups are compared, work environment characteristics must be measured as well.

Most studies do not measure ability or personality (e.g., Cobb, et al., 2009; Hudson & deGast-Kennedy, 2009; Mamo, et al., 2011; Mills & de Araújo, 1999; Schwaab, et al., 2011; Sutcliffe & Alrayes, 2012), or motivation to learn (Cobb, et al., 2009; Mamo, et al., 2011; Mills & de Araújo, 1999; Schwaab, et al., 2011; Sutcliffe & Alrayes, 2012). One quasi-experimental study assigned the first 50 students who arrived to class to the VW condition and everyone who arrived thereafter to the traditional training condition (Cobb, et al., 2009). This is especially troubling because conscientiousness, a personality trait related to punctuality, has been found to predict training performance (Barrick & Mount, 1991), confounding personality with the VW treatment effect. In all of these quasi-experiments, due to incompletely controlled threats to validity, it is difficult to draw substantive conclusions.

Based upon the proposed model, attitudes toward VWs, experience with VWs, and VW climate should always be measured before training begins within each quasi-experimental group being compared. Additionally, personality characteristics and abilities as relevant to the training should be measured as appropriate. If trainers suspect variance in training climate between groups, this should also be assessed. Researchers can then use independent-samples t-test or one-way ANOVA to compare quasi-experimental groups on each of these pre-training characteristics.

5.2 Unmeasured Organizational and Individual Antecedents of Training Performance

Even if quasi-experimentation is not used, studies are embedded within an organizational context that may affect the magnitude of results. Unmeasured trainee characteristics (micro-organizational variables) and work environment variables (macro-organizational variables) may negatively impact the perceived effectiveness of VWs for training. Given that many virtual world studies utilize university student samples, the fact that many ignore work environment characteristics (e.g., Cobb, et al., 2009; Herold, 2009; Hudson & deGast-Kennedy, 2009; Lester & King, 2009; Mamo, et al., 2011; Mills & de Araújo, 1999; Mitchell et al., 2011; Schwaab, et al., 2011; Sutcliffe & Alrayes, 2012) may be unsurprising. However, even within student populations, the university, department or major a student belongs to may promote a culture in support or condemnation of technology in general or VWs in particular.

Many studies make no mention of measuring experience with VWs (e.g., Hudson & deGast-Kennedy, 2009; Mamo et al., 2011; Mills & de Araújo, 1999). Without consideration of VW experience, it is difficult to identify how much variance in an observed difference between groups may be due to virtual world proficiency or lack thereof. One study that did measure experience found that those with virtual world experience were less satisfied with the feedback capabilities of the virtual world than those without (Cobb, Heaney, Corcoran, & Henderson-Begg, 2009), highlighting the relevance of Proposition 2a to reaction outcomes and the value of appropriate consideration of this variable. Most studies of VW-based training lack measures for attitudes toward virtual worlds as well (e.g., Cobb, et al., 2009; Mamo, et al., 2011; Mills & de Araújo, 1999; Schwaab et al., 2011; Sutcliffe & Alrayes, 2012).

In the proposed model, we can see both the proximal and distal effects of both macro- and micro-organizational characteristics on reactions to training, learning outcomes, and behavioral change. Critically, this implies that the expected effect of a negative climate toward VWs would be to decrease the effectiveness of the VW-based training *but not* the effectiveness of traditional training, skewing conclusions if left unmeasured. One study found participants generally disliked training in the virtual world (Herold, 2009), but the characteristics described here measured pre-training could have helped to explain this result; they may have been biased to report this before ever participating. Researchers and practitioners should check for low means in attitudes toward VWs and VW climate before conducting

training. If means are low (i.e. below neutral), training to improve attitudes or climate should be conducted before any VW-based training commences.

5.3 Limited or Missing VW Navigational Training

Skill at navigating a VW is relatively uncommon. If VW experience of trainees is low, a training program on VW navigation must be used to address this, or reactions and learning are likely to suffer, as implied by Proposition 2a. Some studies have no formal virtual world training (e.g., Lester & King, 2009; Mills & de Araújo, 1999). Although these studies found equal performance between VW and traditional conditions, VW navigational training may have enabled learners to utilize the virtual world more effectively, decreasing their cognitive load during training and enabling them to surpass the performance of the traditional trainees.

Simply providing *some* navigational training is not necessarily sufficient. Many studies reported that the VW-based training was not more effective than traditional training despite training (e.g., Heinrichs, Youngblood, Harter, Kusumoto, & Dev, 2010; Herold, 2009; Mamo, et al., 2011; Sutcliffe & Alrayes, 2012). Further, several reported that despite providing some training on the virtual world, participants reported dissatisfaction with the virtual world rating it “not useful” (Heinrichs, et al., 2010) or ineffective and difficult to use (Mamo, et al., 2011). One study reported many groups had difficulty navigating the virtual world even after receiving 20 minutes of VW navigational training, with one group becoming so frustrated they gave up on completing the training task (Mamo, et al., 2011). Researchers should not assume that a brief training accompanied by written instructions is adequate as one study noted that many trainees continued to struggle with basic skills such as moving their avatars and using the chat interface after such basic training (Herold, 2009). A substantial amount of time and effort should be devoted to VW navigational training by training designers to minimize the risk of such catastrophic failures. It also highlights the high priority need for research on the amount and type of navigational training needed to ensure sufficient skill among trainees. This has important consequences not only for reactions but also training performance and ultimate transfer.

In the present model, experience with VWs is theorized as a moderator of the relationship between training design and reactions/learning. Thus, VW navigational training should be implemented with the goal of increasing this experience. VW designers should carefully consider which VW-related skills are central to their VW-based training programs and design navigational training to improve these skills. If simple navigation is all that is required, this should be trained; if more complex skills are needed (e.g. chat interfaces, voice controls), these should be trained as well. Most critically, trainees should not begin VW-based training without the trainer first verifying that the trainees have the requisite skills to successfully navigate VWs.

5.4 Incomplete Exploration of the Outcome Domain

Learning and reaction outcome measures are common in training research in general, and measurement of behavior and transfer is much less common (Van Buren & Erskine, 2002). VW-based training is no exception to this trend with many studies lacking behavioral (e.g., Cobb, et al., 2009; Herold, 2009; Lester & King, 2009; Mamo, et al., 2011; Mills & de Araújo, 1999; Schwaab, et al., 2011; Sutcliffe & Alrayes, 2012) or transfer outcomes (e.g. Cobb, et al., 2009; Herold, 2009; Hudson & deGast-Kennedy, 2009; Lester & King, 2009; Mamo, et al., 2011; Mills & de Araújo, 1999; Schwaab, et al., 2011; Sutcliffe & Alrayes, 2012). The case could be made that virtual world researchers face many of the same obstacles of other training researchers and as result have many of the same criterion measurement problems. However, this is especially detrimental to a field seeking to revolutionize

training with new technology that has a considerable learning curve for both the trainee and trainer. One of the theoretical benefits to VW-based training is that better representation of complex skills enables better learning of those skills. If the only measure of learning in a study on VW-based training is a paper-and-pencil learning test, this advantage cannot be demonstrated. Crucially, the current research literature does not contain any empirical examinations of transfer resulting from VW-based training; we currently lack published evidence to conclude, “Skills learned in VWs can be used on the job.” This is a critical research need.

As suggested by the proposed model, Kirkpatrick’s (1976) training outcomes should always be assessed to address these concerns: reactions to training, cognitive and skill-based outcomes, behavioral change on the job, and macro-organizational outcomes.

6. Conclusions

Based upon the integrative model presented here, negative trainee attitudes toward VWs and a lack of experience with VWs potentially have two detrimental effects: 1) they diminish the effect of good training design on reactions and learning and 2) they diminish the effect of learning from a VW on meaningful organizational results. Negative VW-related climate decreases motivation to learn, which in turn decreases training outcomes. By explicitly considering these individual differences and taking steps to ready trainees before VW-based training, both academics and practitioners can more accurately evaluate the results of VW-based training.

We have presented four major categories of common research- and practice-oriented situations where the model may change conclusions meaningfully as a result of these effects. Given these categories, we present five practical recommendations for training practitioners and researchers:

1. All four outcomes of interest should be assessed if feasible: reactions, learning, behavioral change, and organizational results. This will provide a complete picture of the effect of any training program (see Kirkpatrick, 1976).
2. Attitudes toward VWs, experience with VWs, and VW climate should be measured before training begins, and preferably before training is designed. This will give the training designer or researcher perspective on what might influence the effectiveness of their training program before even beginning. If learners have negative attitudes toward VWs, training targeted at improving those attitudes (Campbell & Kuncel, 2002) should be implemented before VW-based training begins. Without doing so, the designer risks the presence of an unmeasured moderation effect that could undermine her success. Even if attitudes are neutral or somewhat positive, attitude training may improve outcomes further.
3. If comparing VW-based and traditional training across multiple groups, researchers should take care to measure and compare personality (see Barrick & Mount, 1991), ability (see Hunter, 1986; Ree & Earles, 1991), and motivation to learn (see Colquitt, LePine & Noe, 2000) between groups. Independent-samples *t*-tests or one-way ANOVA can be used for this purpose.
4. Practitioners and researchers should consider the organizational context to ensure that no macro-organizational antecedents (like VW climate) are impacting observed effectiveness. Measure these well ahead of training, if feasible. For a review of climate and culture, see Reichers and Schneider (1990).
5. Training designers should provide sufficient VW navigational training such that trainees report they are comfortable navigating and communicating in VWs *before* training begins. A VW experience measure or a focus group can be used for this purpose.

In summary, the integrative model developed here is recommended as an evaluative framework for practitioners and researchers alike. It provides additional explanatory power for training in VW settings not provided by any current model of training evaluation. It extends training evaluation theory generally by updating the Baldwin and Ford (1988) model to reflect recent evidence by Colquitt and colleagues (2000). It also contextualizes theory to the VW environment, providing specific recommendations to ensure accuracy in the evaluation of VW-based training. Establishing the value of VWs accurately and completely should be our top priority.

References

- Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review*, 3, 385-416. doi: 10.1177/1534484304270820
- American Society for Training and Development [ASTD]. (2011). State of the industry, 2011: ASTD's annual review of workplace learning and development data. Alexandria, VA: American Society for Training and Development. Retrieved from <http://store.astd.org/Default.aspx?tabid=167&ProductId=22697>
- Atinc, G., Simmering, M. J., & Kroll, M. J. (2012). Control variable use and reporting in macro and micro management research. *Organizational Research Methods*, 15, 57-74.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41, 63-105. doi: 10.1111/j.1744-6570.1988.tb00632.x
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1-26. doi: 10.1111/j.1744-6570.1991.tb00688.x
- Bedford, C., Birkedal, R., Erhard, J., Graff, J., Hempel, C., Minde, B. ... York, J. (2006). Second Life as an educational environment: A student perspective. In D. Livingstone & J. Kemp (Eds.), *Proceedings of the Second Life Education Workshop at the Second Life Community Convention* (pp. 25-26). University of Paisley, UK. Retrieved from <http://www.simteach.com/SLCC06/slcc2006-proceedings.pdf>.
- Bessière, K., Ellis, J. B., & Kellogg, W. A. (2009). Acquiring a professional "second life": Problems and prospects for the use of virtual worlds in business. *CHI EA 09*, 2883-2898.
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36, 1065-1105. doi: 10.1177/0149206309352880
- Brown, K. G. (2005). An examination of the structure and nomological network of trainee reactions: A closer look at "smile sheets". *Journal of Applied Psychology*, 90, 991-1001.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Burke, L. A. and Hutchins, H. M. (2007). Training transfer: an integrative literature review. *Human Resource Development Review*, 6, 263-296.
- Campbell, J. P. & Kuncel, N. R. (2002). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil (Eds.), *Handbook of industrial, work, and organizational psychology*, vol. 1 (pp. 278-312). Thousand Oaks, CA: SAGE.
- Cheal, C. (2009). Student perceptions of a course taught in Second Life. *Innovate*, 5(5). Retrieved from http://www.innovateonline.info/pdf/vol5_issue5/Student_Perceptions_of_a_Course-Taught_in_Second_Life.pdf
- Cobb, S., Heaney, R., Corcoran, O., & Henderson-Begg, S. (2009). The learning gains and student perceptions of a Second Life virtual lab. *Bioscience Education*, 13.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707. doi: 10.1037/0021-9010.85.5.678

- Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation (4th ed.)*. Belmont, CA US: Wadsworth/Thomson Learning.
- Gronstedt, A. (2007). Second Life produces real training results: The 3-D Web is slowly becoming part of the training industry. *T+D*, 61(8), 44–49.
- Grossman, R. and Salas, E. (2011) The transfer of training: what really matters. *International Journal of Training and Development*, 15, 103–120.
- Hatch, S. (2007). VWs, real meetings. *Corporate Meetings & Incentives*, 26, 12-17.
- Heinrichs, W. L., Youngblood, P., Harter, P., Kusumoto, L., & Dev, P. (2010). Training healthcare personnel for mass-casualty incidents in a virtual emergency department: VED II. *Prehospital and Disaster Medicine*, 25, 424-432.
- Herold, D. K. (2009). Virtual education: Teaching media studies in second life.
- Hudson, K., & deGast-Kennedy, K. (2009). Canadian border simulation at Loyalist College.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340–362.
- Kirkpatrick, D. L. (1976). Evaluation. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (pp. 301-319). New York: McGraw-Hill.
- Kraiger, K. (2008). Transforming our models of learning and development: Web-based instruction as enabler of third-generation instruction. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 454-467.
- 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. doi: 10.1037/0021-9010.78.2.311
- Knowles, M. S., Holton, E. F., III & Swanson, R. A. (2005). *The adult learner: The definitive classic in adult education and human resource development*. Woburn: MA: Butterworth-Heinemann.
- Landers, R. N. & Callan, R. C. (2011). Casual social games as serious games: The psychology of gamification in undergraduate education and employee training. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious Games and Edutainment Applications* (pp. 399-424). Surrey, UK: Springer.
- Landers, R. N. & Goldberg, A. S. (in press). Online social media in the workplace: A conversation with employees. In M. D. Covert & L. F. Thompson (Eds.), *Psychology of Workplace Technology*. New York: Routledge Academic.
- Lee, J. & Kim, S. H. (2010). Analysis on the research trend about the use, research methods, and domains of VWs. *Journal of the Korean Society of Educational Technology*, 26, 163-183.
- Lester, P. M., & King, C. M. (2009). Analog vs. digital instruction and learning: Teaching within first and Second Life environments. *Journal of Computer-Mediated Communication*, 14, 457-483. doi: 10.1111/j.1083-6101.2009.01449.x
- Mamo, M., Namuth-Covert, D., Nugent, G., Phillips, L., Sandall, L., Kettler, T., & McCallister, D. (2011). Avatars go to class: A virtual environment soil science activity. *Journal of Natural Resources and Life Science Education*, 40, 114-121.

- Mania, K., Wooldridge, D., Coxon, M., & Robinson, A. (2006). The effects of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 12, 396–404.
- Mills, S., & de Araújo, M. M. T. (1999). Learning through virtual reality: A preliminary investigation. *Interacting with Computers*, 11, 453–462. doi: 10.1016/s0953-5438(98)00061-7
- Mitchell, S., Heyden, R., Heyden, N., Schroy, P., Andrew, S., Sadikova, E., et al. (2011). A pilot study of motivational interviewing training in a virtual world. *Journal of Medical Internet Research*, 13, e77.
- Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, 39, 497–523. doi: 10.1111/j.1744-6570.1986.tb00950.x
- Reichers A.E. & Schneider, B. (1990). Climate and culture: an evolution of constructs. In B. Schneider (Ed.), *Organizational climate and culture* (pp. 5-39). San Francisco, CA: Jossey-Bass.
- Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, 44, 321–332.
- Reeves, B., Malone, T. W. and O'Driscoll, T. (2008). Leadership's online labs. *Harvard Business Review*, 86, 58–66.
- Schwaab, J., Kman, N., Nagel, R., Bahner, D., Martin, D. R., Khandelwal, S., et al. (2011). Using second life virtual simulation environment for mock oral emergency medicine examination. *Academic Emergency Medicine*, 18, 559–562. doi: 10.1111/j.1553-2712.2011.01064.x
- Shen, J., & Eder, L. B. (2009). Intentions to use virtual worlds for education. *Journal of Information Systems Education*, 20, 225–233.
- Spector, P., & Brannick, M. (2010). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, 14, 287–305. DOI: [10.1177/1094428110369842](https://doi.org/10.1177/1094428110369842)
- Sutcliffe, A., & Alrayes, A. (2012). Investigating user experience in Second Life for collaborative learning. *International Journal of Human-Computer Studies*, 70, 508–525. doi: 10.1016/j.ijhcs.2012.01.005
- Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, 12, 257–285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, 4, 295–312.
- Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology*, 76, 759–769. doi: 10.1037/0021-9010.76.6.759
- Van Buren, M. E., & Erskine, W. (2002). The 2002 ASTD state of the industry report. Alexandria, VA: American Society of Training and Development.
- Velada, R., Caetano, A., Michel, J. W., Lyons, B. D. and Kavanagh, M. J. (2007). The effects of training design, individual characteristics and work environment on transfer of training. *International Journal of Training and Development*, 11, 282–294.

- Wyld, D. C. (2008), World of workcraft: Educating and training digital natives. *Cutter IT Journal*, 21(9), 26-31.
- Wyld, D. C. (2010). A *Second Life* for organizations?: Managing the new, VW. *Management Research Review*, 33, 529-562.