Live or Virtual Military Training? Developing a Decision Algorithm

Christina K. Curnow
Arthur Paddock
ICF International
Christina.Curnow@icfi.com
Arthur.Paddock@icfi.com

Robert A. Wisher
Independent Consultant
Robert.Wisher@gmail.com
Frank C. DiGiovanni
Carl Rosengrant
Office of the Secretary of Defense
frank.digiovanni@osd.mil
carlton.rosengrant@osd.mil

ABSTRACT

This paper describes the development of a decision algorithm for determining what military tasks can be taught virtually (e.g., simulator, advanced distributed learning) and which tasks should only be taught in classroom or field environment (live). The decision algorithm, based on a DoD study, addressed both individual and collective tasks across the military Services. The goal was to develop a user-friendly system to aid military training developers in making 'first-cut' decisions about training delivery methods, specifically live or virtual. To develop the algorithm, we first examined thousands of military training tasks, reviewed the literature on training tasks and developed a rating system to categorize tasks. The categorization scheme resulted in a variety of task classes with each class encompassing common training characteristics (e.g., level of interactivity or availability of feedback). We conducted an extensive review of the research literature and developed rating factors, which formed the basis of the live vs. virtual decision model. We then drew a random sample of 302 military tasks, categorized the tasks and then applied the rating factors to each task category. Next, using the rating factors we developed a decision algorithm for determining whether each class of tasks can be adequately trained using virtual technologies (costs withstanding) or whether it would be necessary to train the task in a live application. The algorithm is based on a variety of elements from established, peer-reviewed research, current technology, and current military practices. Finally, we applied the algorithm to the task categories developed earlier in the project and conducted an initial validation of the algorithm with training developers. In addition to describing the development and validation process, we will solicit feedback and comments from audience members for consideration during further development, validation, and refinement of the algorithm.

ABOUT THE AUTHORS

Dr. Christina K. Curnow is the Director of the Workforce Research Center and a Vice President at ICF International. She has over 18 years of experience conducting research and evaluations related to military training, leadership and distance education. Dr. Curnow has published peer-reviewed journals articles and book chapters and presented at national and international conferences. She has had the opportunity to conduct research within the Army, Air Force, Marine Corps and Joint Forces. Dr. Curnow holds a Ph.D. in Industrial and Organizational Psychology from the George Washington University.

Arthur Paddock is a Senior Associate with ICF International in the Applied Organizational Research Group. His work focuses on applied research projects within training, personnel selection, competency development and organizational assessment. He has directed organizational analysis projects for the Defense Logistics Agency (DLA) and the Air Force Research Laboratory (AFRL). Mr. Paddock received his M.S. in Industrial-Organizational Psychology from the University of Baltimore in 2001.

Dr. Robert Wisher is an Independent Consultant to ICF International. Bob served a distinguished career of 32 years as a Research Psychologist for the Department of Defense followed by two years as a Research Professor at the Naval Postgraduate School. He received a B.S. degree in Mathematics from Purdue University and a Ph. D. in Cognitive Psychology from the University of California, San Diego.

Frank C. DiGiovanni serves as the Director, Training Readiness and Strategy, Office of the Deputy Assistant Secretary of Defense (Readiness). His responsibilities include policy and oversight of military training readiness and capability modernization. He leads the Department's $4.3B Combatant Commander Exercise and Engagement and Training Transformation, the sustainment of military training ranges, the

development of Live, Virtual and Constructive Training Standards and Architectures, the Advanced Distributed Learning Initiative, the creation of a "virtual world" training capability, and ensures training is properly incorporated into major acquisition programs. He also serves as a senior DoD training member on the Modeling and Simulation Steering Committee and collaborates with interagency partners to develop training strategy and policy to ensure Government civilians and Service members are better prepared to conduct reconstruction and stabilization operations. He oversees efforts and policies associated with sustaining access to DoD's land, air and sea training space and for developing policy, strategic communication and the research agenda associated with energy infrastructure and its impact on the ability of the Department to conduct readiness training activities.

Carl Rosengrant serves as the Associate Director for Training Technology as a member of Training Readiness and Strategy, Office of the Deputy Assistant Secretary of Defense (Readiness) staff. He is responsible for development of policy and oversight of training architectures and standards for integrating live, virtual and constructive environments supporting the Department of Defense's training community. Additionally he is charged with championing the Virtual World Framework, an innovative approach to distributing training content via collaborative virtual spaces.

Live or Virtual Military Training? Developing a Decision Algorithm

Christina K. Curnow
Arthur F. Paddock
ICF International
Christina.Curnow@icfi.com
Arthur.Paddock@icfi.com

Robert A. Wisher
Independent Consultant
Robert.Wisher@gmail.com
Frank C. DiGiovanni
Carl Rosengrant
Office of the Secretary of Defense
frank.digiovanni@osd.mil
carlton.rosengrant@osd.mil

INTRODUCTION

This paper describes the construction of a decision-making framework to designate whether military tasks *can* (not necessarily *should*) be trained entirely through virtual methods (e.g., simulator, advanced distributed learning) or need to be taught through traditional live methods (e.g., instructor-led classroom, field site with actual equipment). The framework essentially offers a first-cut, categorical estimation, rendered as a simple radar chart, intended for use by training developers or policy makers working in collaboration with a subject matter expert. The intention is to offer planners a tool that proposes whether to consider further virtual training as an alternative. The study was sponsored by the Office of the Under Secretary of Defense (Personnel & Readiness).

We developed separate frameworks for individual and collective tasks. Within each framework, we propose thresholds that separate virtual from live methods that are sensitive to the current state of proven, off-the-shelf training technology. The thresholds can be adjusted as new technologies become available. The framework is also sensitive to psychological factors established in the literature on individual and team training, which was the emphasis of the study and is the primary focus of this paper.

A review of the classification criteria selected for each framework is presented, first for individual and then for collective tasks. We then detail our assumptions and provide a tally of how several hundred military tasks place in the categorization scheme. We describe the thresholds that divide virtual from live methods, offer examples of the radar charts that reflect the live-virtual decision framework and provide a tangible example.

CLASSIFICATION FACTORS AND CRITERIA FOR INDIVIDUAL TASKS

Although there are numerous task taxonomies and categorization schemes already designed for various purposes, our interest was in isolating those factors that are sensitive to the live versus virtual training issue. (Note: Our use of the terms live and virtual is associated with the live, virtual, constructive taxonomy used in modeling and simulation but applied here to any training delivery, rather than the sense of a virtual classroom common in the distance or online learning communities.) We conducted a review of the research literature to identify factors and criteria upon which to categorize tasks for this purpose. We searched a wide range of task categorization methodologies and sought to isolate specific criteria capable of differentiating tasks solely on the basis of whether they can be trained through a virtual instructional method (i.e, exclusively technology based). We determined that no single task classification model satisfied our objective.

The review identified a set of categorization factors and criteria that can be used to group individual tasks into discrete categories that underpins a live-virtual decision framework. We reviewed models from the following technical perspectives:

- Bloom's Taxonomy – domains such as cognitive, affective, psychomotor
- Levels of Analysis – branch, occupational series, duty position
- Time and Motion Analysis – examination of work systems
- Worker Functions (Functional Job Analysis) – Data, people, things
- Task Characteristics – time spent, difficulty to learn, perishability, importance, frequency, criticality, standardization.
- Position Analysis Questionnaire – information input, mental processes, work

output, interpersonal activities, work situation and job context
- Cognitive Task Analysis – the cognitive structures and processes associated with task performance
- Instructional Requirements that Limit Media Selection – sensory mode, conditional knowledge, synchronous feedback
- Level of Interaction – communication for exploration or for teambuilding
- Perishability and Task Retention Models – how quickly will a particular skill or knowledge be forgotten?

Integration of Technical Perspectives

Our approach examined existing schema on task classification and selected specific criteria from each that differentiates tasks, principally with respect to whether they *can* be taught in a live or virtual environment. Of the criteria that merited further consideration, we looked for similarities and overlap between the constructs advocated from multiple perspectives. The result was a partition incorporating four factors, described below with the categories and criteria.

Domain Factor, categorical scale

We recommend using the *domains* (cognitive, psychomotor, and affective) from Bloom's taxonomy with the added category of "procedural" that fits many military tasks. Although many tasks can be described as a procedure, some include a dominant psychomotor or a critical cognitive component that overrides the routine nature of step-by-step execution that we count as procedural. In the scheme, then, the four categories are:

1. Procedural—routine step-by-step, limited cognitive complexity or psychomotor activity
2. Cognitive—knowledge and development of intellectual skills
3. Psychomotor—involving physical movement, motor skills, or perceptual and physical coordination
4. Affective—involving emotions, motivation, and attitudes

Interaction/Fidelity Factor, ordinal scale

This factor relates to the criteria of data, people, or things derived from functional job analysis as well as the interpersonal activities category of the position analysis questionnaire. We recommend using four categories for this factor:

1. One-way interaction with data or things, low fidelity requirements
2. Two-way interaction with data or things, moderate fidelity requirements
3. Two-way interaction with people, moderate fidelity requirements
4. Two-way interaction, high fidelity requirements

Learning Complexity Factor, ordinal scale

This factor refers to how complicated a task is to learn and how difficult it is to maintain. To determine learning complexity, we recommend using multiple considerations that can be integrated into a single complexity factor. The first two considerations are derived from the work of Rose, Czarnolewski, Gragg, Austin, & Ford (1985) on skill retention. The next consideration relates to mental requirements discussed by Rose et al., Bloom's (1956) levels, and cognitive task analysis criteria. We call this factor *learning complexity*. The mental requirement category is based on Bloom's taxonomy, resulting in five discrete categories with criteria ranging from 'consistently highly complex' to 'not complex at all.'

Task Certainty or Feedback, ordinal scale.

Finally, *task certainty* is the extent to which a task has built in feedback, such that an individual knows when he/she has successfully completed the task without feedback from an instructor. This criterion is based on Rose et al. (1985) and Clark, Bewley, and O'Neil (2006), and has three levels of feedback:

1. Built in/synchronous
2. Sometimes available/Sometimes delayed
3. Never available or very delayed

The four factors were used to establish task classes by assigning a numerical category to each permutation of each domain, much like the Dewey decimal system, where each task receives a numerical assignment for each of the four categories. Applying these criteria to an individual task would result in a four-tuple sequence that identifies a particular class of tasks. For example, marksmanship is a psychomotor skill (category 3), requires interacting with a weapon (category 2), is occasionally complex (category 2), and has built in certainty about whether it has been done correctly (category 1). Therefore, this task would be part of category designated 3.2.2.1 tuple. The total combination of factors results in 240 possible classes of tasks.

CLASSIFICATION FACTORS AND CRITERIA FOR COLLECTIVE TASKS

Psychological research on team performance spans more than 50 years, with a voluminous literature of thousands of studies reporting on the processes and factors that underlie team effectiveness, or what mediates the relationship between team inputs and outcomes. Numerous studies suggest that performance on a collective task can be predicted, to a certain extent, from individual capabilities. But not all performance can be accounted for, and in many cases most cannot be accounted for by simply combining the performance on individual tasks. When substantial interaction between individuals is required, for example, the relationship is greatly diminished. Training of teamwork, demands of synchronous activities, and communication cues come into play.

Taxonomies to describe teams, and team or collective tasks, are plentiful. They tend to focus on particular aspects of team composition and performance, such as selection, internal dynamics, leadership, performance, and numerous other variables. This emphasis on team characteristics is consistent with literature that endorses competency modeling as an analytical technique. Competency modeling involves focusing on the employee characteristics required for effective job performance, rather than focusing on the characteristics of the job itself (Alliger, Beard, Bennett, Colegrove, & Garrity, 2007). The competency-based approach to training can be applied to both individuals and teams training for situations ranging from relatively simple to highly complex (Colegrove, Rowe, Alliger, Garrity, & Bennett, 2009).

The interest in the use of technology, however, has been investigated more as a set of tools used by team members, rather than as a preferred method of training the collective effectiveness of the team. There are numerous categorization schemes for collective tasks, however our interest was in isolating those with factors that address the choice between virtual versus live training.

The literature review further identified a set of categorization criteria that can be used to group tasks into discrete collective task categories that lead to the construction of a decision algorithm. We reviewed models from the following technical perspectives:

- Team performance, input-process-output models (McGrath, 1984)
- Team composition (Dyer, 1984)
- Teamwork and team processes (Bennett, Alliger & Colgrove, 2009)
- Temporal dynamics (Marks, Mathieu, & Zacarro, 2001)
- Environmental factors – e.g., equipment, environment, safety, non-verbal cues

Our model for categorizing collective tasks drew from a chapter in the *Handbook of Industrial/Organizational Psychology* (Cannon-Bowers and Bowers, 2011), which synthesizes many recent models and taxonomies of teams and team tasks. Much of the framework for the decision-making model is rooted in the syntheses presented in that chapter, with added refinements drawn from related articles. The goal was to develop a taxonomic scheme that parallels the taxonomy developed for individual tasks.

To determine an appropriate model for task classification for this study, we examined potential factors identified in the literature review. Of the criteria that had merit for further consideration from the live versus virtual training issue, we looked for similarities and overlap. This led to four factors described in the following section.

FACTORS AND CRITERIA FOR THE COLLECTIVE MODEL

Described here is a summary of the key elements from the research literature that led to the development of four factors. Each is described along with its categories and criteria.

Domain Factor, categorical scale

This factor addresses the nature of the team in terms of what they need to accomplish for a specific task. There is no single, universally agreed on taxonomy of teams. For our purposes, we reduced the classifications to three categories, focusing on the outcomes of team performance in generic terms, stemming from the influential early work by McGrath (1984), which led to many variations of a general input-process-output model. The three categories and criteria are:

Category 1 - Project/Development
Members of this team category are typically involved with planning, analysis of alternatives, and so forth. They likely need to collaborate on project work. An "output" or product may be complex and unique, such as a mission analysis, a course of action, or a piece of software.

Category 2 - Action and Negotiation
Action and negotiation production teams are highly skilled specialists who must cooperate in brief performance events. For our purposes, the main outcome is a decision or recommendation rather than a formal document.

Category 3 - Production and Service
Production and service teams work together in a physical environment where the use of equipment, movement of assets, or reactions to tangible conditions (e.g., terrain) influence performance. This category can include construction teams, assembly line work, or field activities of small units.

Teamwork Training Factor, ordinal scale

Collective tasks can engage more than the knowledge and skills of individuals, (such as teamwork, communication, and physical activities) and may depend on coordinated performance that is not necessarily trained at the individual level. The training of teamwork skills is distinguished from the training of individual skills. Prerequisite capabilities of individual members are essential for successful team training (Dyer, 1984).

The categorization assumes that individuals are proficient on tasks performed in isolation, so teams rather than individuals are the basic unit of analysis. This factor concerns the development of roles and interaction patterns among members of the teams. It consolidates the supporting competencies that underlie successful performance of a mission essential competency used in the Air Force (Bennett, Alliger, & Colegrove, 2009), such as situational awareness, multi-tasking, and internal teamwork. For our purposes, the model simply recognizes teamwork training as a factor with three rating categories (Low (1), Medium (2), and High (3)), indicated by degree to which collective task training emphasizes teamwork.

Synchronous Activity Factor, ordinal scale

This factor concerns the degree to which teams are required to coordinate their actions in order to perform their collective task successfully. These are also known as team processes. This factor relates to the supporting competency of external teamwork, or knowing when, how and to whom to handoff tasks and accept handoff of tasks. A number of taxonomies have been previously proposed as organizers. The model proposed by Marks, Mathieu, & Zaccaro (2001) has the strongest technical support in statistically fitting team processes to team performance. We generalize this temporal dynamic consideration in the collective categorization scheme as overall synchronous activity and the extent to which coordination and task dependencies are present, with three rating categories (Low (1), Medium (2), and High (3)).

Environmental Conditions Factor, ordinal scales

This factor addresses issues that are relevant to instructional delivery, rather than to team processes and performance. These issues have not generally been included in taxonomies of team performance, but they are important for the purposes of the present study. The environmental factor includes four subcategories: The actual equipment subfactor asks whether use of actual equipment is needed, versus using a virtual representation and is rated in three categories (not needed (1), preferred (2), essential (3)). The special environment subfactor addresses whether certain conditions are necessary for collective training, such as darkness or background noise and is rated in three categories (not needed (1), preferred (2), essential (3)). The non-verbal subfactor relates to the synchronous factor but asks directly of the presence of cues that can be seen or otherwise sensed but not heard and is rated in three categories (none (1), occasional (2), frequent (3)). Finally the multi-motoric subfactor seeks to rate whether two or more members of the team must engage simultaneous strength or dexterity in performing an action and is rated in three categories (none (1), occasional (2), frequent (3)).

Decision-Making Framework

The concern here is essentially whether a task is suited for virtual training method. Factors such as costs, infrastructure, courseware maintenance, etc. are not addressed, but obviously would come into play after the first-cut estimation offered in this decision-making framework is made.

Assumptions

Two overarching assumptions were (1) all tasks can be trained through live training methods unless a task is entirely virtual at the outset, and (2) the decision-making methodology requires minimal instruction for trainers or subject matter experts, so we sought to include the most important factors rather than all possible factors. Therefore, we chose a simplified approach with a high, but imperfect, level of accurate prediction rather than a complex approach that may have greater refinement and technical accuracy but is beyond everyday usage. This tradeoff mainly occurred in determining the number of task factors and categories to use in classifying tasks. A separate assumption was that the virtual technology contemplated is currently in use and commercially available, not in an R&D stage, prototype form, or concept formation stage. Specific assumptions were on two topics:

Individual Task Assumptions

A. The model assumes that tasks are trained to a level sufficient for proficiency, as recognized by the Service, using the training method selected;
B. The model does not account for a blended learning approach. Therefore, if a virtual method is selected, then the model assumes that the entire task *can* be trained virtually, so blended learning falls into the live side of the dichotomy;
C. Tasks deemed appropriate for virtual *can* be wholly taught through virtual technology, with no live instructor input other than for administrative and technical procedures;
D. The final certification of task performance can occur either through virtual or live testing, depending on military Service regulations and preferences.

Collective Tasks Assumptions

A. Individuals and subgroups are proficient in all prerequisite individual and subgroup tasks;
B. The model assumes that tasks are trained to a level sufficient for collective proficiency, as recognized by the Service, using the training method selected (i.e., live or virtual);
C. The current model does not account for a blended learning approach. Therefore, if a virtual method is selected, then the model assumes that the entire task can be trained virtually, so blended learning techniques fall into the live side of this dichotomy;
D. When deemed acceptable for virtual training, the collective task is wholly taught through virtual technology, with no live instructor input other than a human-in-the-loop for administrative and technical procedures;
E. The virtual technology contemplated is currently in use and commercially available (not in an R&D stage, concept formation etc.);
F. The size of the collective, or group, is between 5 and 24. The recommendation from the decision-making framework may hold for larger or smaller groups, but with reduced certainty as to its validity.

DEVELOPMENT OF A MILITARY TASK DATABASE

We developed a two-pronged approach to developing a military task database – one for identifying individual tasks and one for identifying collective tasks. The process for developing the database included:

- Identify Individual Tasks by Military Service
- Sample Military Occupations
- Acquire Military Occupational Task Lists
- Acquire Common/Mandatory Tasks Lists
- Identify Collective Tasks by Military Service and Joint
- Identify Universal Task Lists (UTL) and mission essential task lists (METL)
- Explore the use of Joint Tasks

Upon determining the individual and collective tasks to be included in the task inventory, we designed and compiled a database of these tasks.

Initial Application of Individual Task Categories

We identified individual tasks by sampling from military occupational task lists and common or mandatory task lists. From this pool of thousands of tasks, 200 individual tasks were drawn from the Army (61%), Navy (14%) and Marine Corps (26%) and then analyzed according to the classification scheme. These are preliminary data that largely represent the findings from the Army and Marine Corps. The sample was stratified and representative of the types of individual tasks that used the most frequently. In addition, we added tasks to ensure coverage across task types such as the affective domain which tended to be less common.

Tables 1 through 4 provide information about what Service the tasks are from and how they were distributed among the task classification categories.

Table 1. Ratings for Domain Factor

Categories	Percentage
Procedural	38%
Cognitive	34%
Psychomotor	25%
Affective	4%

Table 2. Ratings for Interaction/Fidelity Factor

Categories	Percentage
Low	10%
Medium	39%
Medium/High	32%
High	20%

Table 3. Ratings for Learning Complexity Factor

Categories	Percentage
Not complex	11%
Complex at times, but usually not complex	40%
Moderately complex	34%
Varying between moderately complex and high complexity	13%
Consistently highly complex	3%

Table 4. Ratings for Task Certainty/Feedback

Categories	Percentage
Built in/synchronous	59%
Sometimes available, sometimes delayed	37%
Never available or very delayed	5%

Individual task ratings are represented by 77 types of combined rating categories, of the 240 possible combinations.

Decision-Making Framework for Individual Tasks

Based on the four factor categorization scheme, the literature review of media selection methods, and the set of assumptions outlined above, the recommendation regarding whether a task qualifies to be trained through virtual technologies can be viewed as a decision threshold based on the pattern of ratings. The mapping of where each of the possible 240 combinations of ratings (4 x 4 x 5 x 3) fit into the decision framework is not always purely dichotomous (live vs. virtual). In general, tasks that rated lower on each factor are candidates for instruction through virtual training and those that rate high on each factor are candidates for live training, but there are gray areas. For example, there may be certain psychomotor tasks that have moderate interaction/fidelity ratings that may or may not be suited for virtual training, depending on complexity and task certainty/feedback ratings.

The Radar chart in Figure 1, a multivariate plot of factor rating values, presents a notional view of the decision-making framework, shown as a border inside or outside the LV (Live Virtual) Threshold line. Basically, individual tasks that fall inside the box outlined by the dashed blue line are strong candidates for virtual training. Those that appear outside the blue box probably need to be trained through a live training method.

These notional cutoffs are as follows:

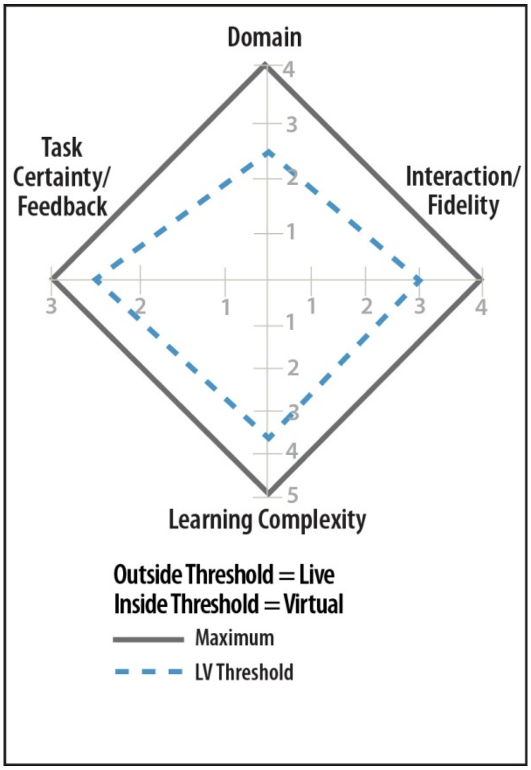
- If Domain is 2.5 or greater, and Interaction/Fidelity is greater than 3, and Learning complexity is greater than 3.5 and Task Certainty is greater than 2.5, then the recommendation is for live training.
- If Domain is 2.4 or less, interaction/fidelity is 3 or less, learning complexity is 3.5 or less, and task certainty is 2.5 or less, then the recommendation is for virtual training.

In addition to these notional cutoffs, we have developed additional cutoffs to be evaluated during the validation of the model. These are not depicted in the radar graph, but are applied in our analyses. These include the following:

- For procedural tasks: If interaction is 1-2, complexity 1-4 and task certainty is 1-3, then the recommendation is for virtual training. Then for all other tasks classes that are in the procedural category, the recommendation is that the tasks are "potentially virtual, but require additional consideration."

- For cognitive tasks: If interaction is 1-2, complexity 1-4 and task certainty is 1-3, then the recommendation is for virtual training.
- For psychomotor tasks – all psychomotor tasks need to be trained live except when fidelity requirements are low (1), in which case this category is "potentially virtual, but requires additional consideration."
- For affective tasks – All affective tasks should be trained live, unless fidelity requirements are low and learning complexity is 2 or less and task certainty is 1 or 2.

Figure 1. RADAR Decision Framework for Individual Tasks.[1]



Initial Application of Collective Task Categories

A total of 102 collective tasks; from the Army, Marine Corps and Air Force; were collected and rated. The data reported below are from collective tasks in the following occupational areas: Field Artillery, Infantry, Military Police, Chemical Biological Radiation Nuclear, Combat Engineer, E-3 Mission Crew, Corrections, Transportation, Personnel & Administration, and Finance. A breakout of collective task by service indicates 53% Army, 12% Air Force, and 35% Marine Corps.

Collective task ratings are represented by 62 sets of combined rating categories. For the Domain factor, Project/Development accounted for 9% of tasks, action/negotiation 30%, and production/service 61%. For the ratings of Teamwork Training, the low category accounted for 22% of tasks, medium 64%, and high 15%. For the ratings for synchronous activity the low category accounted for 21%, medium 58%, and high 22%. Table 5 presents the breakout for tasks in the Environmental factor, with four subfactors.

Table 5. Ratings for Environmental Factor & 4 Subfactors

Actual Equipment Subfactor	
Categories	**Percentage**
Not needed	36%
Preferred	44%
Essential	20%
Special Environment Subfactor	
Categories	**Percentage**
Not needed	61%
Preferred	29%
Essential	10%
Non-verbal Cues Subfactor	
Categories	**Percentage**
None	59%
Occasional	37%
Frequent	4%
Multi-motoric Activity Subfactor	
Categories	**Percentage**
None	60%
Occasional	37%
Frequent	3%

[1] For purposes of graphical representation, the category values for the individual task ratings were transformed to 5-point scales.

Collective Task Decision Framework

Based on the seven factor categorization scheme for collective tasks, the literature review of media selection methods, and the set of assumptions outlined above, the recommendation regarding whether a task qualifies to be trained through virtual technologies can be viewed as a decision threshold based on the pattern of ratings. The mapping of where each of the possible 2,187 combinations of ratings (3 x 3 x 3 x 3 x 3 x 3 x 3) fit into the decision framework is not completely dichotomous (live vs. virtual). In general, tasks that rated lower on each factor are candidates for virtual training and those that rate high on each factor are candidates for live training, but there are gray areas.

Following the decision framework for individual task, the framework for collective tasks is presented as a Radar chart with a similar design logic: those factor ratings that lead to a point inside the innermost polygon, defined by the LV Threshold, are candidates for a virtual training method while those outside this polygon are candidates to be trained through a live method.

In the Radar chart in Figure 2, a multivariate plot of factor rating values, presents a notional view of the decision-making framework, shown as a border inside or outside the LV (Live Virtual) Threshold line. Basically, individual tasks that fall inside the box defined by the dashed line are strong candidates for virtual training. Those that appear outside the blue box probably need to be trained through a live training method.

These notional cutoffs are as follows:

- If Domain is greater than 2.25, and Team work Training is greater than 2.25, and Synchronous is greater than 2.25 and Actual Equipment is greater than 2.25 and Special Environment is greater than 2.25 and Nonverbal cues is greater than1.75 and Multi-motoric is greater than 1.75, then the recommendation is to train in a live environment.
- If Domain is less than or equal to 2.25, and Team work Training is less than or equal to 2.25, and Synchronous is less than or equal to 2.25 and Actual Equipment is less than or equal to 2.25 and Special Environment is less than or equal to 2.25 and Nonverbal cues is less than or equal to 1.75 and Multi-motoric is less than or equal to 1.75, then the recommendation is to train in a virtual environment

In addition to these notional cutoffs, we have developed additional cutoffs to be evaluated during the validation of the model. These include the following:

- Anytime that Multi-motoric activity is frequent (3) then the recommendation is to train in a live environment.
- Anytime that Non-verbal cues are frequent (3) then the recommendation is to train the task in a live environment.
- Anytime that a special environment is essential (3) then the recommendation is to train in a live environment.
- Anytime that actual equipment is essential (3) then the recommendation is to train in a live environment.
- Any time Production/Service (3) is the domain then the recommendation is to train in a live environment.
- Anytime Teamwork training factor is high (3) then the recommendation is to train in a live environment.
- If the domain is Project/Development (1), then the recommendation is to train in a virtual environment, unless the cutoffs for Multi-motoric, non verbal, special environment of actual equipment specified above are met.

Individual Task Example

An example of individual task ratings, as displayed in the radar decision framework, is listed below. The task, shown in Figure 3, displays the ratings of an Army task for Radar Repairers (Monitor Bench Stock Operations) in comparison to the LV threshold. As shown, based on the assumption of the framework, this task *can* likely be trained in a virtual environment.

Figure 2. RADAR Decision Framework for Collective Tasks.

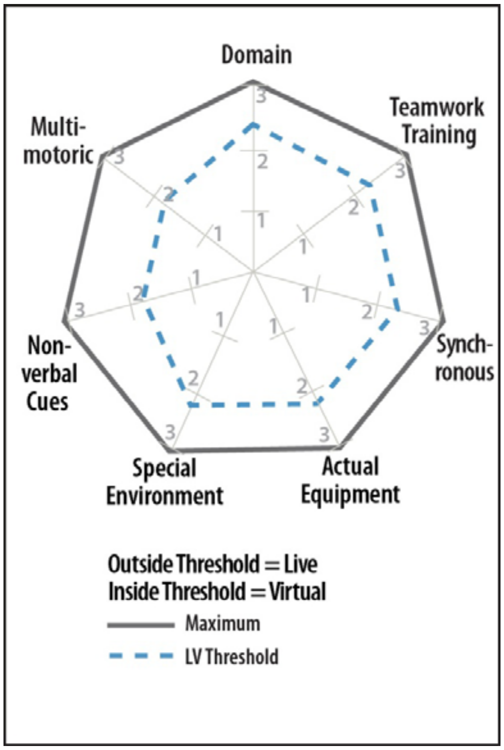
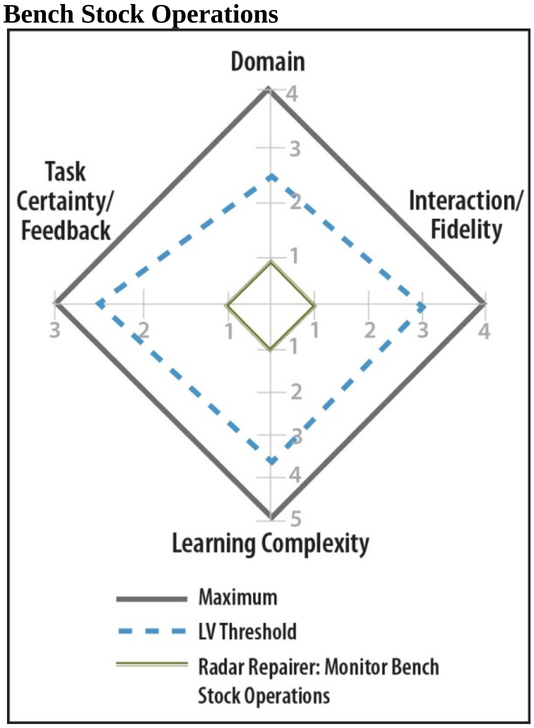


Figure 3: Radar Decision Framework for Individual Task – Radar Repairer: Monitor Bench Stock Operations



Inter-rater Agreement of Factor Ratings

The strategy for rating tasks across factors is dependent on development of consensus ratings from multiple raters. Therefore, dependence on agreement of raters during preliminary steps is mitigated. In order to gain additional insight into factors that influence consensus ratings, we examined the agreement among individual raters at this preliminary stage. Appropriate inter-rater agreement statistics were calculated for ratings of initial samples of tasks used to test the model.

For the individual task model, 154 tasks had individual rater data from two raters. Results of analysis of inter-rater agreement varied across factor scales based on data type (i.e., categorical, ordinal, scale) and varying scale levels (e.g., 3-, 4-, and 5-level scales). Examples include the Interaction/Fidelity and Learning Complexity factors. The average measure of interclass correlations (i.e., assuming average ratings across raters) for Interaction Fidelity is 0.6 and Learning Complexity is 0.7, supporting the strategy to use consensus ratings

Next Steps and Future Applications

Based on the tasks that we collected and rated to date, and our review of the literature, we have preliminary decision frameworks for both individual and collective tasks. The next step in this process will be to validate the model and get feedback from potential users. Based on the outcome, we will plan to revise and refine the decision framework to make it more useful for the intended user audience.

REFERENCES

Alliger, G., Beard, R., Bennett, W., Colegrove, C., & Garrity, M. (2007). *Understanding mission essential competencies as a work analysis method*. AFRL-HE-AZ-TR-2007-0034, Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate.

Bennett, W., Alliger, G., & Colegrove, C. (2009). Mission essential competency. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay, Inc.

Cannon-Bowers, J.A., & Bowers, C. (2011). Team development and functioning. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization* (pp. 597-650). Washington, DC: American Psychological Association.

Clark, R. E., Bewley, W. L., & O'Neil, H. (2006). Heuristics for selecting distance or classroom settings for courses. In H. O'Neil & R. Perez (Eds.), *Web-based learning: Theory, research and practice* (pp. 133–142). Mahwah, NJ: Lawrence Erlbaum Associates.

Colegrove, C. M., Rowe, L. J., Alliger, G. M., Garrity, M. J., & Bennett, W. (2009). *Defining the training mix – sorties, sims, and distributed mission operations.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Dyer, J. (1984). Team research and team training: A state-of-the art review. In F.A. Muckler (Ed.), *Human factors review: 1984* (pp.285-323). Santa Monica, CA: Human Factors Society.

Marks, M.A., Mathieu, J.E., & Zaccaro, S.J. (2001). A Temporally based framework and taxonomy of team processes. *Academy of Management Review, 26,* 356-376.

McGrath, J. E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice

Rose, A., Czarnolewski, M., Gragg, F., Austin, S., & Ford, P. (1985). *Acquisition and retention of soldiering skills* (Final Report 671). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.