The Arrival of Analytic Evaluation Strategies for Training Systems

Dr. Regan J. Patrick
Program Manager, KDAM ATARS
CAE USA Mission Solutions
regan.patrick@caemsi.com

ABSTRACT

As education and training programs improve their ability to collect increasingly large volumes of information on learner performance, a need has emerged to reexamine legacy evaluation strategies and processes. 'Big Data' sources are enabling training system designers to leverage analytics and Artificial Intelligence capabilities to create individualized learning approaches more closely aligned to student needs. Armed with an increased understanding of how students achieve competency, program developers and leaders require insight into how they can improve both individual and program achievement. The challenge for instructional designers becomes how to 'design-in' feedback mechanisms that contribute to intelligent decision-making. Feedback data from a wide variety of sources can now measure not only the effectiveness of student accomplishment, but also the efficiency and affordability of all enterprise processes aligned with evolving expectations of training system customers and consumers.

Instructional designers need to know how to access individual and enterprise performance data to validate current training system design and shape future education system constructs. We know technology generates data, but what data should feed advanced AI algorithms to support operational and resource decisions at the local and program levels? How can emerging education and learning technologies integrate with analytic strategies to optimize Training Return on Investment (T-ROI) and future value propositions? How can program managers know they are meeting both student and end user requirements?

This paper examines emerging processes and technology applications behind evaluation strategies designed to provide meaningful input to complex training programs. It offers approaches to help instructional designers, managers, and instructors apply feedback to shape long-term learning behaviors, refine syllabus requirements, and improve overall system performance aligned with end user expectations. The author recommends effective ways to examine data on how students achieve their learning objectives, and describes a framework to help training managers know what should be measured and how much, from whom and how often; questions fundamental for training system optimization today and tomorrow.

ABOUT THE AUTHOR

Dr. Regan J. Patrick is the CAE USA Mission Solutions Program Manager for the Kirtland AFB, Davis-Monthan AFB, Joint Base Andrews, and Moody AFB Aircrew Training and Rehearsal System program (KDAM ATARS), based in Tampa, FL. He leads a team of over 300 employees providing academic instruction and simulator training to 1,500 aircrew students per year at four sites, serving U.S. Air Force Special Operations, Combat Rescue, DV Airlift, and Nuclear Surety missions. Dr. Patrick is a retired USAF officer and former HH-60G Pave Hawk helicopter evaluator pilot with over 2,400 hours of flight time and multiple contingency and combat deployments worldwide. His military assignments included tours as Deputy Operations Group Commander, Training Squadron Commander, and Acquisitions Program Manager. He holds a Doctor of Education in Organizational Change and Leadership from the University of Southern California, a Master of Aeronautical Science Degree from Embry-Riddle Aeronautical University, and a Bachelor of Arts in Economics from San Diego State University. A former Air Education and Training Command Master Instructor, Dr. Patrick is a graduate of the USAF Air War College, Department of Defense Advanced Instrument School, and the Joint Air Commander School. He served on the adjunct faculty at Embry-Riddle Aeronautical University's Worldwide Campus for 11 years, and he is an IRONMAN triathlete.

The Arrival of Analytic Evaluation Strategies for Training Systems

Dr. Regan J. Patrick
Program Manager, KDAM ATARS
CAE USA Mission Solutions
regan.patrick@caemsi.com

INTRODUCTION

Data science has expanded our understanding of how students achieve competence in a training system and provided insight into system-wide changes that can improve individual and program-level achievement. Technology-fed approaches to learning are enabling instructional designers to create individualized education experiences better aligned to student needs and enhance training system responsiveness and effectiveness. As our ability to collect, interpret, and apply performance data improves, the need to reexamine legacy evaluation strategies has grown as well. How can instructional designers 'design-in' feedback mechanisms that work together to improve decision-making processes individually and at scale?

This paper seeks to explore and conceptualize an approach for building a training system evaluation strategy that leverages the power of data analytics and Artificial Intelligence (AI) to simultaneously improve both individual and enterprise performance. It begins by establishing definitions for key concepts, and examining current practices in evaluation and assessment, and the formation of evaluation strategies. The relationship between evaluation and performance improvement for students and training systems is examined through the application of feedback pathways and framework. Integrating data science approaches such as analytics and AI to evaluation strategies is then considered using an example from the U.S. Air Force. Finally, integration and execution approaches are considered and discussed. A conceptual example is also provided, showing how an analytic evaluation strategy can be applied to a current training system and the potential benefits it offers.

UNDERSTANDING EVALUATION AND ITS LIMITATIONS

The importance of evaluation in education has been clearly established in the literature and in practice. It is an essential ingredient in any effort to close performance gaps or improve performance, and the only process that provides an objective view of progress (Clark & Estes, 2004). Evaluations are measurements used to determine the degree to which knowledge, skills, and other attributes have been acquired (Shute & Ventura, 2013). There is a direct correlation between quality instruction and quality assessment; value-added assessments of learning are the product of a clear and specific target, controlling for sources of interference unrelated to achievement, that sufficiently sample student performance in a manner representative of the eventual performance domain (Stiggins, 1991).

The New World Kirkpatrick Model of Training Evaluation, a modern extension of the oldest, most often cited, and most widely used evaluation framework provides three reasons for conducting evaluations of student assessment: improving the program, maximizing transfer of learning, and demonstrating the value of training to the organization (Kirkpatrick & Kirkpatrick, 2016). While the benefits of evaluation are clear, a cohesive evaluation strategy to manage the results can provide essential input to inform critical program decisions in an uncertain future.

Definitions and Delineations

Before we begin any discussion on analytic evaluation strategies, some definitions are in order. For the purposes of this paper, a *training system* is defined as all resources necessary to achieve the goals of a learning program, including personnel, courseware, hardware, instructional methodology, and evaluation approaches. It represents a complex interaction of goal setting, resource allocation, time, and commitment on the part of instructional designers, administrators, and educators. Training systems can be small, serving a limited number of students learning relatively simple tasks to large, with hundreds of students learning complex, interdependent performance-based skills. They may provide graduates to a single organization or dozens.

Professor and author Michael Watkins recently provided a definition of business strategy which applies to training system evaluation strategies as well. He points to a set of "guiding principles" that must be clearly understood by all members of an organization, used to generate a "pattern of decision making" that defines future actions and prioritizes resources to achieve goals (Watkins, 2007). In an education context, these guiding principles represent end user requirements that must be clearly understood by all training stakeholders. They will describe success or failure at every step of the learning process, shape and reorder individual learning approaches, and inform decisions throughout the lifecycle of the training system. Clearly defining the problem and goals is the first, and most important step in any strategy development effort.

To improve the odds of successfully accomplishing program goals in a highly dynamic, rapidly changing learning environment, training system designers are must carefully consider how evaluations support critical decisions at all levels across the enterprise. Progressive evaluation strategies represent an evolutionary step away from industrial or "factory" models of learning, dominant throughout the 20th century, where teaching is simply transmission and learning happens at the end (Abbott, 2019). All students are taught the same way, in the same timeframe, and evaluated in exactly the same manner.

Future adaptive evaluation methodologies will synthesize high-volume feedback pathways from diverse, data-rich sources to inform change in both individual and program learning outcomes in near real-time. Shifting from the industrial to digital age will require new ways of measuring, analyzing, and applying information in learner-focused, as opposed to compliance-centric, instructional experiences (Crisp & Bonk, 2018).

Digital tools to apply evaluation feedback abound for developers in the education space. *Artificial Intelligence* is a prediction technology used to assist decision makers confront uncertainty in the face of information overload (Agrawal, Gans, & Goldfarb, 2018). *Machine Learning* (ML), a subset of AI is used to optimize processes and sharpen resource allocation; learning from experience and improving performance as it grows (Bini, 2018). *Big Data* sources are defined as high-volume, high-velocity, and/or high-variety information assets of such size and complexity that computer processing is required to effectively and efficiently sort, analyze, and interpret the data (Gartner, n.d.). *Data analytics* uses large data sets, statistical and quantitative analysis, explanatory and predictive models, and fact-based management decisions to find useful patterns in data (Davenport & Harris, 2007). As the volume and speed of information in a training system increases, a strategy is needed to effectively harness these digital tools.

As a starting point for any evaluation strategy, training system leaders must carefully define the meaning of the word "success" for individuals and the program itself (Moore, Lockee, & Burton, 2002). It begins with a requirement; a tangible outcome defined by the end user. This critical feedback (or 'feed forward') establishes the "guiding principles" for the entire educational process. The Kirkpatrick evaluation model advocates developing results-oriented outcomes in collaboration with all stakeholders during planning, and embedding mechanisms for collecting data from the beginning (Moreau, 2017).

Categories of Evaluation and Assessment

Instruction and assessment have a symbiotic relationship; they're inexorably linked and interactive, and a variety of different types of assessments should be threaded throughout the course of instruction (Abbott, 2019). For the purposes of this paper, *evaluation* can be defined as an attempt to discern whether a system does what it was designed to do efficiently and effectively, while an *assessment* refers to a measurement of learner performance before or after instruction has occurred (Moore et al., 2002). While assessments can be included as a part of an evaluation, and an overall evaluation strategy, the two aren't necessarily synonymous or interchangeable. Assessments and evaluations represent performance feedback related to the goals established by the training system end user.

In 1967, Professor Michael Scriven first introduced the general categories of formative and summative evaluations. *Formative evaluation* involves an ongoing, planned process to collect "assessment-elicited evidence" of a learner's status to adjust instructional procedures, which students can also use to adjust their current learning tactics (Popham, 2008). It provides important feedback to learners on ways to improve their performance, and to instructional designers on how they can improve their processes and products (Gagné, Wager, Golas, Keller, & Russell, 2005). Formative evaluations can be used throughout the entire education experience to help instructors identify struggling students early and know where to apply resources, and to help students identify areas they must focus on (Carnegie Mellon University, 2020). *Summative evaluations* occur at the end of an instructional unit, comparing performance against a standard while attempting to reflect how students will perform in the real world. They represent “pass, remediate, or fail” decisions, determining whether students have achieved a satisfactory level with respect to establish criterion (Gagné et al., 2005). The deliberate application of both approaches can provide valuable insight into both learner and instructional system performance.

DEVELOPING AND DEPLOYING AN EVALUATION STRATEGY

The idea of developing a strategy to guide evaluation processes in educational systems is not new. Indeed, many evaluation models, including the New World Kirkpatrick Model include provisions for strategy development, though they tend to focus primarily on improving individual learner performance. Understanding what should be evaluated and when is the foundation of an effective strategy.

An evaluation strategy helps instructional designers provide clearly-stated and measurable objectives, and apply appropriate instructional strategies to ensure students learn what the system goals and objectives demand with sufficient practice and feedback mechanisms (Moore et al., 2002). This involves building a roadmap to help students achieve learning expectations, and help the training enterprise produce sufficient numbers of students with the desired knowledge, skills, and attitudes to meet end user requirements – which can change quickly and substantively.

Evaluation strategies generally follow three primary objectives – improving *efficiency*, *effectiveness*, and *affordability*. Efficiency describes volume, or the use of resources to achieve throughput objectives per time, instructor, or any other constrained variable. Effectiveness refers to knowledge transfer and behavior modification; an accountability measure to ensure students are learning what they need to know in a consistent, repeatable manner to meet the expectations of the end user. Affordability means accomplishing education goals within time or financial constraints.

The goal of an evaluation strategy is to create a robust process to capture, analyze, and act on inputs (or feedback) from key stakeholders. To explore this, let’s consider a scenario involving a technical training program seeking to optimize individual and enterprise performance – initial qualification training for military helicopter pilots. This is a complex endeavor, involving task certification on hundreds of individual skill sets with a fixed production requirement (number of graduates per year). All students arrive at this program with the same basic flying skills, though they will require individualized and adaptive learning approaches to produce the deep learning required to ensure subject mastery. The standards are unyielding, and graduates must complete the program and transition to their gaining units on-time and mission ready. As new systems, tactics, and procedures are introduced throughout the lifecycle of the aircraft, the training system must be able to quickly adapt and adjust. In many cases across the U.S. military, programs like this are structured and conducted under a production-centric, industrial model of operation. A carefully thought-out evaluation strategy is essential to help training system leaders meet these complex goals and shift the training methodology into the 21st century.

How Evaluation Can Fall Short

Evaluating student and enterprise performance comes with risk. There is often a tendency to evaluate programs, instructors, and learners in terms of process rather than product, which will not provide useful information in closing performance gaps (Stiggins, 1991). Outdated or ineffective models of assessment, which are often summative and decontextualized of individual learner performance can lead to inaccurate conclusions (Abbott, 2019). Confounding variables in an evaluation model and causal chains of evidence can produce uneven or inconclusive results (Moreau, 2017). Evaluation strategies that fail to support deep learning, or the acquisition of complex competencies by relying on ‘snapshot in time’ approaches to judge student learning can lead to false assumptions and inaccurate analysis (Shute & Ventura, 2013). Poorly structured assessments can create an, “overwhelming focus on criteria compliance and award achievement” that produces students who are more dependent on instructors and mentors rather than less dependent (Torrance, 2003). Sound evaluation strategies require care and consideration to effectively support learning goals.

If a student successfully passes a lesson or course an assumption can be made that knowledge transfer has occurred, objectives are understood and met, and the program was effective in achieving its goals. However, asking students to simply memorize a process or procedure for a test does not guarantee they will be able to achieve the desired level of performance on the job (Clark & Estes, 2004). Exclusive reliance on summative evaluations risks creating a sense of "fluency illusion", where an evaluation that differs significantly from practice creates a false assumption that a learner's ability to remember and apply skills in practice will translate to performance environments (Abbott, 2019).

Inadequately conceived evaluation strategies can also lure students into adopting a "pump and dump" attitude towards learning; cramming as many rote facts into their short-term memory as possible, dumping them onto a 'representative sampling' check or summative exam, and subsequentially forgetting substantial portions of what they 'learned' once they leave the classroom. In progressive skills development programs such as flight training or neurosurgery, this stunted metacognitive approach is a harbinger of long-term struggle for the student and poor training system performance. Unfortunately, it is also all too common in many learning programs today.

USING FEEDBACK TO IMPROVE PERFORMANCE: A SYMBIOTIC RELATIONSHIP

Evaluation strategies draw their strength and effectiveness from valid, reliable, timely, and consistent feedback from key stakeholders. Clearly defining the expectations, components, and pathways feedback will follow is a further critical step in evaluation strategy design and execution. Instructional designers must provide students with clear definitions and dimensions of feedback to empower them as informed consumers of their own learning experience (Crisp & Bonk, 2018). In an individual context, *feedback* can be defined as, "specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance" (Van De Ridder, Stokking, McGaghie, & ten Cate, 2008). It can be more broadly defined as information a system uses to make adjustments in reaching a goal (Ende, 1983). Feedback is essentially a measurement in time, representing a cycle that connects inputs and outputs.

Feedback informs evaluation, and evaluation informs process. Carefully designed feedback pathways are used to make essential program adjustments while highlighting the dissonance between the intended result and the actual result (Ende, 1983). Recently, Artificial Intelligence and analytics have created new approaches for finding and harvesting feedback data, expanding the frequency, speed, and capabilities of modern evaluation strategies.

Feedback Pathways for Training System Evaluations

A training system feedback map can be helpful to outline critical information pathways and understand how formative and summative data can structurally flow within the instructional model (see Figure 1.). This can also contextualize and inform prediction and decision-making models in the Instructional Systems Design (ISD) process.

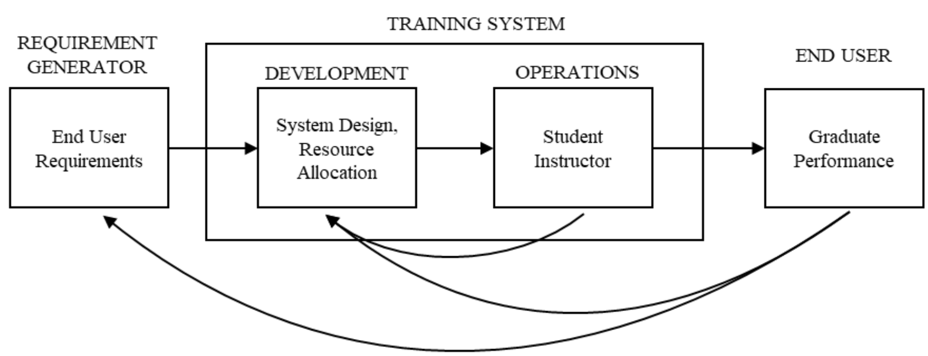

Figure 1. Evaluation Feedback Pathways in a Training System

In this simple illustration, the 'Requirement Generator,' representing the goals and objectives of the end user, assembles and communicates the learning expectations to the Training System, which is composed of two primary nodes – Development and Operations. Once instructional designers understand requirements, they apply a variety of tools to assemble a training framework and syllabus, and allocate resources. Operations leaders then develop schedules, staff the program, and align resources to start the program. As students receive training and are assessed, detailed performance feedback is collected from a variety of internal and external sources. The 'End User' may be the same organization as the 'Requirement Generator,' though they serve distinctly different functions. In our helicopter pilot training scenario, 'Operations' represents the training squadron, while the 'End User' is the gaining operational squadron.

Figure 1. shows how feedback pathways flow forwards and backwards, creating circular nets between nodes ('Operations' and 'Development,' for example). 'End User' requirements shape instructional design and resourcing. Design guides operations of the training system, including expectations of student performance. Student performance is continually assessed and evaluated within the 'Operations' node, which informs program and design adaptations. As the student completes training and moves on, the gaining organization provides feedback to the 'Development' node to assess whether the training system is meeting production quality and quantity requirements.

Internal and External Feedback

Opportunities to evaluate performance occur throughout the entire process, both within the training system and external to it. Students are measured against syllabus requirements, while programs are evaluated against quality and quantity demands. The notional diagram in Figure 1. outlines how effective feedback pathways can be designed in from the beginning of the program to ensure critical data reliably reaches key stakeholders.

Internal feedback within the training system is primarily used to assess individual student performance, while external feedback examines the training system's ability to meet the educational requirements. It also provides a conduit for program-level changes, such as an original requirement that must be deleted or modified, or a new requirement that must be added. These processes exist in many learning programs with varying levels of effectiveness. Data science can speed this process up considerably, while increasing the quality and quantity of information used to make these programmatic changes.

Of particular note, feedback from the end user is often an afterthought or overlooked as unnecessary, even though it is an especially critical component in any evaluation strategy. Training system designers can build in robust critique pathways to include follow-up questionnaires for obtaining this important data (Stiggins, 1991). Post-graduation surveys and interviews, which can be provided in a number of personal or digital formats are effective methods for collecting useful programmatic outcome information (Moore et al., 2002). Interviews of graduates, their managers, and everyone affected by the learner's new knowledge and performance can provide valuable evaluation metrics for determining the success of the training program (Clark & Estes, 2004). Analytical approaches to collecting and interpreting this individual and program feedback data is empowering training systems in new and important ways.

Feedback to Improve Individual Performance

To assist in changing individual learner performance, feedback must be frequent, high quality, and honest. The positive effects of timely, relevant feedback in K-12 and post-secondary settings have been validated in numerous studies (Crisp & Bonk, 2018). Quantitatively-defined dimensions of feedback can also enable the effective evaluation of a learner's experience (Dirlam, 2017). Importantly, if useful feedback isn't available or provided, learners are unable to relate their performance to either current or future learning goals (Abbott, 2019). Quality summative and formative feedback that identifies deficiencies and ensures learners achieve required mastery levels before moving on is an essential ingredient in any effective evaluation strategy.

Feedback on individual student performance provides them with substance and context to improve their learning outcomes, and ensure the program is meeting student needs. Measurements of individual learning should be conducted under criterion-referenced assessment; a comparison with the required criteria for mastery rather than by comparison with other learners (Reigeluth, Myers, & Lee, 2016). Sound instructional practice requires instructors to focus on clear expectations with their students, time feedback intervals to ensure student attention, and check frequently to confirm they have a clear understanding of feedback intent and purpose (Stiggins, 1991).

The appropriate intervals for knowledge and performance evaluation is unique to each program. However, frequent competence assessments should be conducted using pre-defined rubrics that specify performance level expectations. This provides both the student and training system leadership with the exact status of learning at defined points along the learning continuum (Voorhees, & Bedford-Voorhees, 2016). Instructional designers and leaders share responsibility for determining how often student feedback should be sought and at what point(s) within the program.

Effective feedback occurs within a process loop similar to the one in Figure 2. Once instruction has concluded, students are called to demonstrate comprehension and performance through an assessment. This generates feedback through an exam score, physical activity, observational data, or any number of different evaluation criteria. Faculty, instructional leaders, and/or developers will make adjustments to future learning approaches to ensure satisfactory progress is achieved before allowing the student to continue. Changes to instruction are then applied, and the cycle continues. Quality feedback from summative and formative sources enables an instructional system to cycle through these steps continuously, providing useful data to learners and organizations to improve their performance.

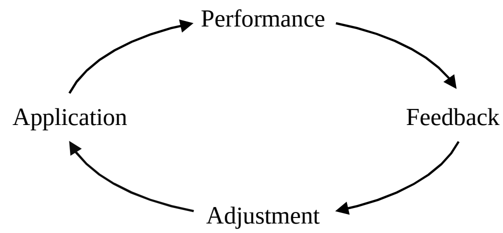

Figure 2. Performance Feedback Loop

A successful example of applying formative and summative evaluation feedback strategies to improve individual learner performance can be seen in the U.S. Air Force's Pilot Training Next, or PTN program (Losey, 2018). Air Education and Training Command has begun testing new approaches to training student pilots by integrating feedback data from education technologies such as AI, Virtual Reality, biometrics, and data analytics to accelerate production, improve learning outcomes, and reduce costs. Innovative assessment and evaluation approaches are used throughout training to support adjustments to instruction. Training is tailored to individual needs, providing personalized attention to help learners grasp, rehearse, and demonstrate complex performance-based tasks in minimal time. The 'Sage on the Stage' is quickly being replaced by the 'Guide on the Side' thanks to creative application of evaluation feedback. PTN has entered its third iteration and is preparing for a service-wide release. The Air Force hopes to use lessons learned from this learner-focused evaluation strategy in other fields of technical training, including medicine and maintenance.

Feedback to Improve Program Performance

In a larger context, training systems must be able to provide compelling evidence that their programs are delivering bottom-line results that contribute to overall mission accomplishment, and that the resources expended went to good use (Kirkpatrick & Kirkpatrick, 2016). To efficiently interpret internal and external sources of feedback, training developers must design feedback pathways into the process from the beginning, not as an afterthought or as time may permit. If training programs can successfully leverage feedback to inform future learning strategies they are better able to answer the most important question of all – how do we know our program is meeting its requirements?

Program evaluation areas can include a wide variety of topics including costs (time or money), graduate performance measured against a standard, quantity and timeliness of graduates, and many others. Training system designers, leaders, and end users own joint responsibility for defining which metrics a program should use to evaluate key areas of efficiency, effectiveness, and affordability.

Regardless of which pedagogical model program designers choose, feedback plays a critical part in successful outcomes (Smith & Dillon, 1999). Feedback data on enterprise performance comes from a variety of sources, both internal and external to the system, and in different forms. Training system leaders must be consistent, insistent, and persistent in obtaining quality feedback from all key stakeholders, especially end users, to ensure an accurate picture of performance is created and maintained throughout the life of a program. Pre-formatted paper or digital surveys, phone calls, emails, or websites can all serve this need. For example, the U.S. Air Force uses an "Aircrew Graduate Evaluation Program," or AGEP as a program performance metric to obtain summative feedback from gaining units on graduate performance in the field.

In their comprehensive Program Evaluation Guide for Public Health Programs, the Center for Disease Control (CDC) has provided general recommendations for program evaluation that can be applied to virtually any training system. Evaluations should serve a useful purpose, be conducted ethically, and produce accurate findings to help leaders make informed decisions and improve overall effectiveness (Self-Study Guide, 2012). The CDC highlights five categories of evaluation questions (Implementation, Effectiveness, Efficiency, Cost-Effectiveness, and Attribution). Collectively, these metrics can help training system leaders shape learning approaches, refine syllabi, inform the implementation of new education strategies, and improve overall system operations.

A conceptual framework can be helpful in illustrating how feedback within an evaluation strategy might be applied to drive training system improvements (Figure 3). This example shows how feedback sources can inform 'change engines' that drive improvements to training system efficiency, effectiveness, and affordability.

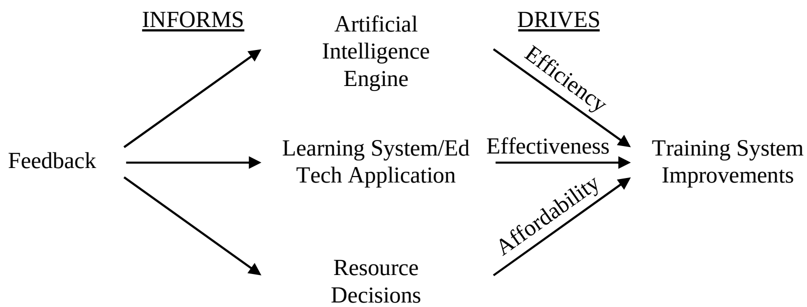

Figure 3. Feedback Pathway to Drive Training System Improvements

Internal and external feedback is collected and assimilated through the evaluation data pathways codified during the instructional design process shown in Figure 1. Feedback data can be used by AI engines to produce predictions leaders can use to drive program efficiencies, reduce cost and time to train, and ultimately improve their training return on investment (T-ROI). Inputs from customers, industry, academia, instructors, and the learners themselves can help instructional designers apply optimized learning methodologies for a given learning task and identify appropriate self-efficacy strategies. They can also assist developers as they weigh the value of alternative education technologies. Feedback plays an important role in resource decisions throughout the lifecycle of a training system by highlighting how and where a program might choose to invest limited resources. It can also shape purchasing and sustainment decisions on equipment, personnel, new instructional media, software and hardware upgrades, and instructor professional development.

Establishing a Feedback Framework

To improve training system effectiveness, requirements analysis within an evaluation strategy should be a constant, ongoing process, informed through AI predictions as well as individual judgement. Establishing a feedback framework helps instructional designers know what to evaluate, how often, how much, and from whom. Table 1. provides a notional example. Actual training systems require detailed, individualized, and contextualized components within each category, tailored to their unique needs and end user requirements.

Table 1. Evaluation Strategy Feedback Framework

WHAT TO EVALUATE (Sources and Components of Feedback):
- Prerequisites and Pre-examinations (prior to training)
- Student Performance (Assessments, Evaluations, Reviews)
- Costs (Time, Resources, Manpower)
- Instructor Performance (Instructor Professional Development, Student Evaluations, Critiques)
- Throughput (Volume and frequency of graduate production)
- Courseware (Lesson Plans, Instructional Media, Handbooks, Supporting Documentation)
- Student and Instructor Guides, Syllabi, Lesson Plans, Course Maps

HOW OFTEN (Intervals):
- Consistently delivery (Frequency, Rate)
- Prior to lesson start
- During performance exercises, rehearsals
- End of Lesson, Phase, Course
- After graduation, at gaining organization (How long after training is completed?)

HOW MUCH (and How):

- Targeted, specific, short, impactful samples
- Formalized, Standardized, and Targeted questions (Rubric-driven)
- Easily Accessed (Mobile, Learning Management System)
- Surveys, Interviews, Training Boards, User Conferences

FROM WHOM (Source):

- Students
- Faculty
- Program Leadership
- Instructional Designers
- End Users

Student assessments provide important feedback to help decision makers know how and when to make critical program adjustments, such as adding or deleting learning requirements, or revising a syllabus (Abbott, 2019). In our helicopter pilot training example, a syllabus of instruction might call for several aircraft flights to achieve mastery of a specific task after ground training and simulation. However, feedback in the form of consistently early performance mastery in flight might indicate that additional flights are either unnecessary or they may be reduced in future syllabus revisions. This can save time and money while expanding program capacity and efficiency, particularly if the process is automated.

INTEGRATING AI AND ANALYTICS IN EVALUATION STRATEGY

Artificial Intelligence can trace its practical beginnings to the 1950s. Though it has historically been constrained by the computational limits of modern microprocessors, recent advances in technology have sparked a renaissance in capability and opportunity. AI engines apply summative and formative feedback data to generate predictions leaders can use to enhance decision accuracy and reduce costs, while improving quality-adjusted prediction outputs (Agrawal et al., 2018). AI has become an important tool in instructional design and performance, though actionable information from AI engines requires data from valid, reliable, and authentic assessments (Stafford, 2019).

Similarly, analytic approaches in data science have begun to unlock the power and promise of ‘Big Data’. The science of Analytics enables training system evaluation strategies to process massive amounts of data, improve decision-making, and explore new areas of capability and performance. This impressive form of Machine Learning uses probabilistic modeling to transform algorithmic problems into prediction problems. In other words, learning from mistakes over time to produce significantly more accurate predictions than regression techniques alone (Agrawal et al., 2018). ‘Big Data’ sources in education have become deeper and more abundant with technology-assisted learning methods such as simulators and interactive courseware. Data analytics and AI are now able to synthesize these complex feedback data streams to recommend quantifiable individual and program-level improvements.

Emerging data analytics and AI capabilities are providing the source code for the “pattern of decision making” Watkins described in strategy formation. Feedback data serves three roles in Artificial Intelligence; it feeds algorithms to produce predictions, it generates an algorithm good enough to “predict in the wild”, and it improves the algorithms’ performance with experience over time (Agrawal et al., 2018). Through analytics, evaluation feedback improves visibility into how learners develop while enhancing an instructor’s ability to be more responsive and adaptive. It also helps instructors recommend new individualized experiences and learning pathways (Abbott, 2019). ML algorithms can uncover learning trajectories, forecast enterprise-level planning issues, and suggest system-wide improvements and changes to strategy (Blake-Plock, 2019). A key strength of AI is its ability to learn continuously through feedback. The more students flowing (or who have flown through) a training system, the better AI will be at understanding trends and making accurate predictions regarding performance improvements, and how to close any performance gaps.

For our helicopter pilot training example, military flight programs produce a substantial amount of student cognitive, metacognitive, and performance-based feedback data. As the PTN example has shown, academic exams, verbal evaluations, checkrides, and instructor observations can be combined with continuous digital performance assessments and metadata collected during academic, simulator, and live flight training to create a continuous and rich source of feedback information. Training squadrons can leverage these expansive data sources to improve individual learning processes, and also connect their learning ecosystems to the requirements of their gaining units. As end user requirements change, program and individual lessons can also quickly adjust through these digital interconnections.

EXECUTING AN ANALYTIC EVALUATION STRATEGY - OPTIMISM AND CAUTION

Training evaluation models are abundant, and evidence of their effectiveness is well established. A recent meta-analysis of education evaluation frameworks examined 14 competing theories of practice, exposing 17 'content categories' (Botek, 2018). Determining which of the models or content categories is most appropriate for a given context is a matter of preference and experience for instructional designers, who can now choose from a rich palette of options for optimizing an analytic evaluation strategy. Successful strategies require knowing what to evaluate and how assessment should be conducted to achieve performance improvement (Clark & Estes, 2004). Evaluation strategies grounded in analytical approaches to data collection and interpretation provide an eclectic and functional system of learning assessment, measuring and processing a vast array of variables from quantitative, qualitative, estimated, and predictive data types (Abbott, 2019).

Defining the Analytic Evaluation Strategy

An analytic evaluation strategy is a plan, or roadmap to integrate large volumes of feedback data from numerous evaluation sources to create near-real time predictive and analysis capability. The output provides actionable information to program leaders to quickly and effectively execute individual and program-level changes. For instance, individual performance data from a student pilot in our helicopter training example can be used to make rapid adjustments to their program of instruction (practice opportunities, drills, reviews, alternate techniques). That same data can also inform program-level changes affecting many students in the same course. Is the courseware or instructional methodology appropriate for this task? Is the time allotted for subject mastery appropriate and at the right time in the training flow? Individual performance data can be combined with data from other students to evaluate overall program performance against end user requirements. Feedback data from an evaluation network flows quickly and deliberately to key stakeholders, powering decision-making processes for individual learners and the enterprise level simultaneously.

Deploying an evaluation strategy fed by expansive internal and external data sources and aided by digital processing capabilities provides program leaders critical information for system-wide improvements. These include a solid cost-benefit analysis, an improved understanding of the motivational impacts of training, and most importantly, undeniable proof that the training program has met the desired learning and production objectives (Clark & Estes, 2004). To improve on their analog predecessors, analytic evaluation strategies synthesize formative and summative assessments that benefit from a wide array of data sources to provide decision recommendations. They also establish an integrated assessment scaffold (the "bones" of instruction) to frame individual lesson assessments (Abbott, 2019).

Integrating data analytics into an evaluation strategy at scale can be accomplished through a four-phase framework (Blake-Plock, 2019). Phase 1 begins with *framing the training problem*, which involves goal setting, needs analysis, and data requirements. Phase 2 involves *data visualization and design* to determine what data sources are best suited to inform performance indicators, including feedback pathways and intervals, as shown in Figure 1. Phase 3 is *architecture development*, where learning-data specifications are established to create clear, domain-based modeling structures used across all levels of the evaluation effort. Phase 4 is *deployment*, when issues of privacy, security, quality, fairness, accountability, transparency, and ethics are also addressed. Ultimately, AI will become better than humans at factoring in complex interactions among different indicators of a training system (Agrawal et al., 2018).

Linking Data Science and Evaluation

An example of a vexing assessment problem affecting large-scale training enterprises is the requirement for a regular, comprehensive Training System Requirements Analysis (TSRA) in U.S. Air Force acquisition and services programs. A TSRA is defined as, "a systematic approach to front-end analysis of a training system based upon an integrated instructional systems development/systems engineering process that develops data items to document the training and preliminary system requirements" (Secretary of the Air Force, 2002). This complex, expensive, and time-consuming analysis process is dependent on many internal and external data sources. Currently, TSRAs require the digital and analog collection of diverse sources of information, organizing and interpreting the data, and producing a final report – a process than can take many months and drive considerable cost. Once an analytic evaluation strategy is established, key stakeholders will have ready access to TSRA-required data on demand and in near real-time. Linking data science and evaluation methodology represents an important step forward in analysis and decision-making capabilities for future learning ecosystems.

Targeting an Analytic Evaluation Strategy on Outcomes

Initially, analytic evaluation strategies require considerable human interaction and participation. AI is a learning technology, and human involvement is necessary to supervise and correct mistakes as the machine learning algorithms and data sources are connected, synthesized, and integrated (Agrawal et al., 2018). Over time and with increasing data sets, AI processes will learn from their early mistakes and reduce their dependency on personal judgement for correction. Eventually, training systems will be able to make adjustments, alter processes affecting outcomes, and anticipate and correct problems at an increasing rate, while gaining speed and precision with every new student. AI-fed evaluation strategies will also provide new and better methods for managing program risks.

Adopting analytical approaches to feedback analysis does not completely remove the need for human participation in training system decisions, however. An analysis of the decision-making process itself will reveal which human activities will diminish in value and importance, and which will increase as a result of enhanced AI capabilities. Artificial Intelligence carries a fundamental limitation about how much it is able to predict preferences, or make a judgement on the value or relevance of given data (Agrawal et al., 2018). Since human beings and data analytics are good at different aspects of prediction, decision making should not be completely automated. Both must understand the limits of the other, and a process should be implemented to capitalize on the strengths of each. AI can provide an initial prediction that humans combine with experience and judgement, a second opinion after the fact, or a path for monitoring. However, the best outcomes will likely be delivered by a synthesis of AI-generated recommendations and human judgement (Agrawal et al., 2018). Combining digital prediction with human judgement helps strategic decision making in training systems become more effective, efficient, and affordable despite varying levels of uncertainty.

For the learner, an analytic evaluation strategy begins with their initial qualification training and carries forward throughout their professional life. An important output from this process is the creation of a student profile of learning, which is an evolutionary move away from the notion of traditional gradebooks. Summative report cards can be replaced with comprehensive and rich electronic portfolios of student work that will follow a learner throughout their career (Moore et al., 2002). These learner databases leverage feedback from past education and training experiences to formulate course and lesson design to meet future learning requirements. Persistent learner profiles form a domain map of individual attainments in knowledge and skills, identify future learning needs, and help programs select appropriate instructional content and methodology (Reigeluth et al., 2016). Performance feedback contributes to a "formative learner profile," or "360° views of learners in real-time," founded on computational learning analytics that empower training system leaders to make quicker, better informed decisions for individuals and the training system at scale (Blake-Plock, 2019).

Warnings for an Analytic Evaluation Strategy

While digital approaches to evaluation strategy provide great promise to future learning ecosystems, they come with warnings to training system designers and leaders. How can instructional designers better define the potential hazards that come with integrating and synthesizing diverse sources of feedback? Unconstrained and misinformed application of technology always carries risk. Failing to offer sufficient interaction between instructors and learners inhibits a student's ability to remain engaged, construct their own knowledge, and develop self-regulation and self-efficacy skills and ability (Abbott, 2018). AI is a prediction machine, and if the machine does not completely understand the decision process used to generate data from feedback, its predictions can fail (Agrawal et al., 2018). AI does not perform well with limited data or uncertainty; it can struggle to extrapolate causality without a tightly controlled and standardized environment to operate in (Agrawal et al., 2018). Privacy concerns, data integrity, information security, and access restrictions can limit availability of data needed to generate important operational recommendations. More study is needed to fully account for concerns regarding the use of analytics in evaluation strategies, but these challenges are clearly solvable.

As training system designers and leaders examine and conceptualize their own approaches to building analytic evaluation strategies, they must also consider how their approaches may change under Competency-Based Education (CBE) constructs in future learning ecosystems. Unlike many current instructor-centric learning approaches, CBE holds performance mastery as the constant and time as the variable. Summative and formative assessment and evaluation methodologies would require adjustment under CBE, though robust feedback pathways would retain their importance. Regardless of which instructional methodologies or learning frameworks are applied, investing time, thought, and resources into the ongoing monitoring of the transfer and persistence of training systems is absolutely necessary for a training system to demonstrate value (Clark & Estes, 2004).

An Analytic Evaluation Strategy in Execution: A Primer for the 21st Century

With these concepts in mind, let’s revisit our helicopter pilot training scenario once more to explore how an analytic evaluation strategy can be successfully applied. It begins with the gaining operational unit(s) clearly communicating knowledge and performance requirements for graduates at the beginning of instructional design; defining what ‘right’ and ‘success’ look like. Feedback pathways are established and formalized between designers, program leaders, instructors, and the gaining units before the first student arrives. Instructional designers assemble an evaluation strategy feedback framework to create formative and summative assessment opportunities designed to gather frequent and timely student knowledge and performance data. This information will be used to improve individual learning experiences through adaptive approaches to instruction.

Analytic models and AI engines within the evaluation strategy will leverage individual and program feedback data of increased scale, complexity, and detail to modify and tailor lessons in near real time. Computer-aided analysis of millions of individual ground and flight performance data points will also ensure common standards are applied and met. These valid and reliable feedback data sources of increasing sophistication and insight are collected through pathways that were mapped at program inception. This information improves program leadership’s understanding of student performance while providing increasingly accurate recommendations for any change that may be required. As new equipment, systems, tactics, and procedures are introduced to the aircraft in the field, training requirements are quickly analyzed, adapted, and updated through feedback pathway models – in some cases, automatically. Information on the training system’s overall performance and capacity is quickly and accurately provided to key stakeholders, including how many students are trained for a given cost, effectiveness of current instructional methodologies, and time to train limitations.

Analytic approaches enable improvement in the responsiveness of the training system to student needs and end user requirements simultaneously and seamlessly. They provide critical information to inform resource decisions and ensure a sufficient number of graduates able to meet the needs of their gaining units, whose requirements will expand and change throughout the lifecycle of the weapon system. They also create a mechanism to better understand and support student needs throughout their professional careers. The means for knowing the training system is meeting the requirements of their customers have been built into the training system from its very design.

CONCLUSION

Evaluation is a crucial part of training system design and operations, and in the accomplishment of individual and program-level educational goals. As education programs evolve in content and complexity, schoolhouses will be challenged to carefully examine how they perform assessments and evaluations, and how they leverage the increasing volume of available feedback data to maximize effectiveness, efficiency, and affordability. A thorough understanding of what ‘success’ looks like will help training system designers and leaders get there. An analytic evaluation strategy built on robust feedback networks can inform and shape both individual and program-wide performance. Analytics and the rise of Artificial Intelligence have provided new tools for training systems to link feedback to predictive and decision-making capabilities, expanding responsiveness and agility to end users requirements and goals.

REFERENCES

Abbott, D. (2019). Assessment and Feedback. In Walcutt, J. J., & Schatz, S. *Modernizing Learning: Building the Future Learning Ecosystem*. (pp.163-180). Advanced Distributed Learning Initiative.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction Machines: the simple economics of artificial intelligence*. Harvard Business Press.

Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty, 33*(8), 2358-2361. doi:10.1016/j.arth.2018.02.067

Blake-Plock, S. (2019). Analytics and Visualization. In Walcutt, J. J., & Schatz, S. *Modernizing Learning: Building the Future Learning Ecosystem*. (pp.163-180). Advanced Distributed Learning Initiative.

Botek, M. (2018). Comparison of Education Evaluation Models. *Calitatea-acces la succes.*

Carnegie Mellon University. (2020). Formative vs Summative Assessment - Eberly Center - Carnegie Mellon University. Retrieved May 25, 2020, from https://www.cmu.edu/teaching/assessment/basics/formative-summative.html

Clark, R. E., Estes, F., Middlebrook, R. H., & Palchesko, A. (2004). Turning research into results: A guide to selecting the right performance solutions. *Performance Improvement, 43*(1), 44-46.

Crisp, E. A., & Bonk, C. J. (2018). Defining the Learner Feedback Experience. *TechTrends, 62*(6), 585-593.

Davenport Thomas, H., & Harris, J. G. (2007). Competing on Analytics: The New Science of Winning.

Dirlam, D. K. (2017). *Teachers, learners, modes of practice: theory and methodology for identifying knowledge development*. Taylor & Francis.

Ende, J. (1983). Feedback in clinical medical education. *Jama, 250*(6), 777-781.

Gagné, R.M., Wager, W.W., Golas, K.C., Keller, J.M., & Russell, J.D. (2005). Principles of instructional design, 5th edition. *Performance Improvement, 44,* 44-46.

Gartner. (n.d.). Information Technology Gartner Glossary. https://www.gartner.com/en/information-technology/glossary/big-data

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation.* Association for Talent Development.

Losey, S. (2018, October 02). The Air Force is revolutionizing the way airmen learn to be aviators. Retrieved May 24, 2020, from https://www.airforcetimes.com/news/your-air-force/2018/09/30/the-air-force-is-revolutionizing-the-way-airmen-learn-to-be-aviators/

Moore, M., Lockee, B., & Burton, J. (2002). Measuring success: Evaluation strategies for distance education. *Educause Quarterly, 25*(1), 20-26.

Moreau, K. A. (2017). Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox?. *Medical teacher, 39*(9), 999-1001.

Popham, W. J. (2008). *Transformative assessment.* ASCD.

Reigeluth, C. M., Myers, R. D. & Lee D. (2016). The Learner-Centered Paradigm of Education. In Reigeluth, C. M., Beatty, B. J., & Myers, R. D. (Eds.). *Instructional-design theories and models, Volume IV: The learner-centered paradigm of education.* (pp.7-32). Routledge.

Secretary of the Air Force. (2002, November 1). *Information for Designers of Instructional Systems* (AFH 36-2235). Department of the Air Force. https://static.e-publishing.af.mil/production/1/af_a1/publication/afh36-2235v3/afh36-2235v3.pdf

Self-Study Guide - Program Evaluation - CDC. (2012, May 11). Retrieved May 25, 2020, from https://www.cdc.gov/eval/guide/index.htm

Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games.* MIT Press.

Smith, P. L., & Dillon, C. L. (1999). Lead article: Comparing distance learning and classroom learning: Conceptual considerations. *American Journal of Distance Education, 13*(2), 6-23.

Stafford, M. (2019). Competency-Based Learning. In Walcutt, J. J., & Schatz, S. *Modernizing Learning: Building the Future Learning Ecosystem*. (pp.243-268). Advanced Distributed Learning Initiative.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, *10*(1), 7-12.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, *14*(3), 281-294.

Van De Ridder, J. M., Stokking, K. M., McGaghie, W. C., & Ten Cate, O. T. J. (2008). What is feedback in clinical education?. *Medical education*, *42*(2), 189-197.

Voorhees, R.A., & Bedford-Voorhees, A. (2016). Principles for Competency-Based Education. In Reigeluth, C. M., Beatty, B. J., & Myers, R. D. (Eds.). *Instructional-design theories and models, Volume IV: The learner-centered paradigm of education.* (pp.7-32). Routledge.

Watkins, M. (2007). Demystifying strategy: The what, who, how, and why. *Harvard Business Review, 10.*