CALL SIGNS

USN AEP SOCIETY

April **2021**

A bi-annual publication of the US Naval Aerospace Experimental Psychology Society

Psychological Measurement

The quantification of human cognitive potential

Uniformed scientists on the forefront of tomorrow's technological battlefields, work to apply state-of-the-art theory to solve the toughest problems

Quantifying the Value of Research

Making strategic investments in future technological dominance require robust methods to measure potential value

The science of human performance has never been more important, nor has its potential been more attainable.

SURVEY DESIGN

Today's methods applied to inform tomorrow's decision. How survey design shapes the research that shapes our futures.

PAGE 4

INFERENTIATION

Answering tough questions often boils down to choosing the right story to tell. Math is no exception

PAGE 9

INVEST IN SCIENCE

Return on investment is a common means of justifying future investments. Quantifying the benefits of research helps make better decisions

PAGE 14

CONTENT

FROM THE USNAEPS PRESIDENT

On behalf of the United States Aerospace Experimental Psychology Society (USNAEPS) Executive Committee (EXCOM), welcome to another issue of Call Signs. This issue focuses on Psychological Measurement. AEPs are uniquely qualified to support the Navy in this area of practice. Recently there has been quite a bit of attention on applying "Metrics that Matter;" this is especially challenging when measuring abstract constructs, but having the ability to operationally define those constructs opens the door to tackling complex problems. Attempting to quantify and measure "Situational Awareness, Training Effectiveness, or Return on Investment" is difficult but with the right skills, statistical tools, clearly articulated assumptions, and operational definitions, such constructs can be tackled and produce actionable data.

With this in mind, I hope you enjoy this issue of *Call Signs!*

CDR Brent Olde, AEP #122

AEPs Past, Present and Future,

Greetings! I am thrilled, honored and humbled to have been selected to serve you as our community's 21st Specialty Leader. As I enter my fifth month in this role, I remain profoundly aware of the high bar set by our outgoing Specialty Leader, CDR Hank Phillips, as well as by the many other Specialty Leaders preceding him. I intend to ensure that our community builds on our past and current successes to not only meet, but exceed, this bar.

Our community has an amazing history of supporting the operational requirements of our Warfighters, delivering new and innovative operational medical capabilities to our Naval Forces — dating back to World War Two. Back then, as part of the "Pensacola Project," we applied our scientific expertise to give Naval Aviation a technical edge. We ensured the best-qualified aviation candidates were selected and trained and solved Human Factors challenges with building, operating and maintaining the newest and most advanced aviation platforms.

Today, as part of the Medical Service Corps, we continue to apply our scientific and technical expertise to solve Human Systems Integration and Operational Medicine challenges in Naval Aviation and across the Navy/Marine Corps. We use our program management skills to transform these solutions into new capabilities delivered to the Fleet. We leverage our analytic abilities to guide development of doctrine, policy, standards and strategy that impact our people and the platforms they operate.

As we look to the future battlespace, with its emphasis on Integrated All-Domain Naval Power, our knowledge, skills and abilities are needed more than ever. Our National Defense Strategy demands that we build and maintain a lethal and ready Joint force that possesses a decisive advantage over any adversary, anywhere, anytime. The Chief of Naval Operations' Navigation Plan and the Surgeon General's Operational Order provide guidance and direction to meet these demands. This includes developing technologies that allow our Warfighters to outmaneuver and outfight our adversaries. It also includes developing technologies that prepare, protect and care for our Warfighters as they engage with our adversaries. AEPs are uniquely positioned to be the link between these two distinct but intertwined technology development approaches.

I recognize that our community, like many other Health Care Scientist communities, faces some unique challenges in the current environment. Having been an AEP for over 20 years – and having served as the Assistant Specialty Leader for three of those years – I know that one of our greatest strengths lies in our deeply rooted sense of Community. AEPs stand by each other!

I plan to leverage this "Unity of Community" to implement a set of goals that will help us manage these challenges. These goals include:

Continue to Recruit the Very Best: The success of our community lies in our collective ability to continue to recruit professionals who are not only brilliant in their areas of scientific expertise, but who also want to apply their expertise to serve our Country and have the drive and potential to grow into future Navy leaders.

Train for Excellence: As a community, we must maintain our expertise on the fundamentals: the behavioral sciences, aerospace and operational medicine and related disciplines. We must also ensure we retain our alignment with broader opportunities, like Defense Acquisition career paths – especially in light of the pending shifts in how the Acquisition community, to which so many of us belong, is trained, certified and managed.

Communicate Transparently: Our community is small, and in many instances AEPs are assigned to "one of one" positions. Under these conditions it is easy for any of us to feel disconnected from our AEP family. I will make sure to maintain consistent and transparent discussions regarding key community-wide actions and opportunities, including establishing the billet slate, recruitment progress, potential for new billets and other key activities that impact us.

As the Surgeon General has noted, we are in a transformational time for military medicine. I am absolutely confident that together we will rise to the challenge of ensuring our Warfighters are trained, equipped and ready to dominate any mission and beat any adversary.

Looking forward!

CAPT Joseph Cohn, AEP #113

Uptatempores se corrovitati doluptur aut fugias magnamus, cusapel ipsantis ulpa dolorum autem que veles dicitibus.Nimus atem dolupta sperchi lluptisci dis suntemolore.

SURVEY DESIGN & MANAGEMENT

Lessons learned from the investigation of complex phenomena in real-world environments

by LCDR Brennan D. Cox

Following publication of the Navy's comprehensive review (CR) on the increase in physiological episodes (PEs) experienced by jet aircrews (2017), faculty of the Naval Postgraduate School's (NPS) Human Systems Integration (HSI) program partnered with NAVAIR and the Physiological Episodes Action Team (PEAT) to conduct a large-scale survey to inform ongoing root cause and corrective action (RCCA) activities. This tasking was consistent with the unconstrained resource approach recommended by the CR, while also aligning with the NPS mission of delivering operationally-relevant research through student and faculty collaboration – in this case, through a directed studies course on survey research methods. The processes and lessons learned from this activity are summarized herein, with an emphasis on survey design and management.

Processes

The NPS Survey Team consisted of five faculty members and three graduate students enrolled in the NPS HSI masters-degree program. During the academic quarter of Fall 2017, we met twice weekly in a classroom setting, initially to review best practices in survey research methods, item writing, and

The Naval Post Graduate School (NPS) offers first-class graduate degree programs in a wide variety of disciplines to eligible U.S. military students, international students, DoD civilian employees, and a limited number of defense contractors. Aerospace Experimental Psychologists are eligible to serve as professors at NPS, and commonly teach courses in Psychology, Statistics, Leadership, Human Factors, and Human-Systems Integration. Photo by Petty Officer 2nd Class Patrick Dionne

Developing a high-quality survey instrument takes dedicated expertise, determined effort, and a lot of patience. Using survey tools to uncover and better understand complex interconnected phonenomena such as physiological episodes in military aviation is an example of how AEPs apply their skills to real-world problems and provide solutions.

scale development, and later to apply this material toward development of the PE Survey. This included a literature review to better understand the PE problem and aircraft systems involved, followed by consultation with the program offices responsible for F/A-18 and EA-18G, and Naval Undergraduate Flight Training Systems.

Each Program Office shared their RCCA fault tree diagrams, which provided top-down graphical representations of the components supporting each aircraft system subject to investigation, as well as their initial questions of interest (nearly 1,000 in total). The Survey Team evaluated the fault trees to identify which sections would best lend themselves to exploration through survey methods (several were eliminated). We then used a content-mapping process to link the initial item list to the remaining fault tree sections, and filled in any gaps with questions developed by the Survey Team. Through this process, six content themes emerged: procedures, flight gear, training, physiological aspects, the cockpit environment, and command culture. It also became evident that separate surveys would be necessary to target the unique experiences of F/A-18 aviators, T-45 aviators, and F/A-18 maintainers.

Next, the Survey Team visited the Program Offices to verify the item-fault tree linkages and to ensure each item adequately addressed the information it was intended to clarify. Items were revised and then used to create pilot testing instruments to evaluate the proposed survey content. NPS students representing each of the three populations (i.e., F/A-18 and T-45 aviators, and F/A-18 maintainers) completed the pilot tests, and subsequently engaged in feedback sessions to comment on the survey content, item wording, and areas in need of revision. Based on their review, the items were again revised and provided to the Program Offices for final approval.

Once all stakeholders agreed to the survey content, the final copies were entered into the private and secure LimeSurvey software tool hosted through NPS. Next, the Survey Team acquired the requisite approvals from the NPS Institutional Review Board (IRB), the Navy Survey Manager, the USMC IRB, and the USMC Survey Manager, as well as email contacts of all potential subjects from the OPNAV and MARCOR staffs (~6500 email addresses). The surveys were launched in January 2018 and remained accessible for 30 days. When the surveys closed, data were downloaded and cleaned, and the analyses began. In total, 1223 (~19%) of those contacted provided usable data. The final report provides detailed analyses corresponding to each of the six content themes and concludes with a list of areas the authors believe warrant attention, recommendations for future work, and notable limitations (NPS, 2018); however, because its contents contains sensitive information, it is not for public release and will not be further discussed in this article.

From start-to-finish, this large-scale DOD survey effort took nearly one year to complete based on the following timeline of events:

- Initial stakeholder meetings and literature review (September 2017).
- Analysis of the RCCA fault trees and initial questions (October 2017).

- Item development and survey refinement (November 2017).
- IRB and survey office approval process (December 2017).
- Data collection (January 2018).
- Analysis and publication of findings (February 2018 – August 2018).

Lessons Learned

From the outset, the Survey Team recognized the PE problem to be an extremely complex, mutli-faceted, system of systems challenge; one that, by its very nature, could not be reduced to a single point of failure; and, given its history of mishaps and public scrutiny, would be sure to generate highly-emotional responding when addressed through an anonymous survey. We were also tasked with meeting multiple objectives. The engineering teams at NAVAIR desired feedback specific to aircraft components (e.g., ECS, OBOGS, and LOX), while our aeromedical partners were more interested in the human experience (e.g., mask discipline; frequency and duration of symptoms). It quickly became clear that designing a single study to address all domains of interest would be an impossible task.

Lesson Learned #1: Understand the strengths and limitations of surveys – and manage expectations accordingly.

This important first lesson is one of acceptance – recognizing what surveys can do despite their limitations and making sure all stakeholders are aware and understanding of the same. Like many psychological measurements, surveys provide a snap-shot in time. The manner in which respondents answer one day may be different than how they might respond at other times; this could be based on situational demands, the sensitivity of the items, and a host of other "human factors." Although this observation may seem obvious to students of the social sciences, it is less likely to resonate with those more accustomed to scientific laws than theories.

For instance, it is well-known that engineers, such as those working at NAVAIR, are well-versed in physics and mathematics, and Newtonian principles relating weight, mass, acceleration, and the transference of energy on the movement and structural integrity of mechanical objects. These phenomena are governed by scientific laws, such that they are universal, absolute, repeatable, and stable. For this reason, they allow for conclusions to be drawn from a particular set of conditions with certainty: if this : then that. The laws of science offer the greatest level of precision for how we understand the world, because, as far as we are able to comprehend, all objects in the universe comply with them. Unfortunately, human behavior is far less predictable, and the measurement of human behavior is far less precise.

Airman Andrew Morse, an aircrew survival equipmentman, verifies the part number on an oxygen mask regulator aboard the aircraft carrier USS Nimitz (CVN 68). The Nimitz Carrier Strike Group is on a routine deployment to the region. Identifying the root causes of physiological episodes is a complex undertaking that extended far beyond the cockpit. (U.S. Navy photo by Petty Officer 3rd Class James Mitchell/ Released)

As a self-report measurement technique, surveys collect information on how individuals perceive the world. They are influenced by recent events, subject to social desirability, and biased by the beliefs, opinions, experiences, and interpretive lens unique to each respondent. Many will recognize this as a problem of measurement error. Of course, error comes in several forms. Students of the hard sciences are likely to be familiar with systematic error, as in the case of an uncalibrated scale that adds a set weight to each object placed upon it. Systematic errors derive from properties of the device or environment in which the measurement is taking place; the key feature is that these errors are systematic, such that they influence the quality of the measurement in a consistent manner. Random error, on the other hand, is just that: random. Random errors inject artificial variability into the outcome of a measurement. In survey research, this type of error is often due to the individual's physical or psychological state at the time of responding, resulting in unintended, unknown, and in many cases uncontrollable influences based on mood, fatigue, motivation, mental condition – or any other infinite artifacts of the human experience.

Surveys are unable to provide an assessment of ground truth. However, they offer many benefits if designed and used appropriately. A survey, for instance, is not the best tool for testing whether an oxygen mask works according to its design specifications. However, a survey can offer insight into the degree to which users perceive the mask to work, which taps into trust. Surveys can also evaluate levels of comfort, ease of use, and perceived value of the device, all of which influence the user's experience. Furthermore, surveys can be made anonymous, which increases the likelihood that respondents will self-disclose information they otherwise wouldn't, such as the frequency and duration they refuse to wear their mask in accordance with policy. These are areas in which surveys offer advantage over other measurement techniques.

Another challenge emerges from the popularity of surveys – people come across them all the time, whether online, on the phone, or in person. We are so often invited to provide feedback, ratings, and evaluations that we may, in many cases, come to believe we are all survey experts. But this, of course, is not the case, and is but one more example of how our personal experiences bias our judgment. This is where true survey experts must educate their stakeholders, by taking time to establish up front what type of information can and cannot be derived from a self-report questionnaire.

For our part, the NPS Survey Team sought to manage expectations from the beginning. In our initial meetings with NAVAIR and the PEAT, throughout all follow-on interactions, and in the opening remarks of our final report, we placed caveats on our role in the RCCA process. Specifically, we stated that the findings from the surveys were unlikely to reveal a "smoking gun" that would solve the PE problem or close out items on the fault trees, but they could be used to inform existing efforts and identify new areas not yet explored.

Lesson Learned #2: Encourage stakeholder participation in the survey development process.

An appropriate extension of the expectation management conversation is to invite stakeholders to participate in the survey development process. With the PE Surveys, this participation occurred naturally as NAVAIR had pre-drafted a list of questions that they wanted in addressed in the surveys. Still, it is recommended that an open dialogue take place throughout the survey development process to ensure the resulting product captures the full intent of the sponsor and to maximize buy-in.

To this end, the NPS Survey Team held frequent phonecons with the NAVAIR engineering team to discuss item content, wording, and fault tree linkages, and visited the Program Offices in person to obtain a final round of feedback. Members of the respondent population were also invited to complete preliminary copies of the surveys and make recommendations for improving acceptability of the instruments; for instance, by incorporating terms and phrases that are well known and used among aviators (and less "sciencey"). Our pilot testing group also helped to ensure the PE Surveys did not come across as offense, accusatory, or insensitive in nature, given the high likelihood that many respondents would have personal experiences with mishaps, injuries, or even loss of life involving PE events.

Lesson Learned #3: Know your question before you write your questions.

An additional offshoot of expectation management involves clarification of purpose – not only of the survey itself (i.e., what you are setting out to accomplish), but also the individual items within (i.e., what information does the item provide and how can it be used). The established purpose should be promulgated by the survey sponsor. It serves as the ultimate guide for how users should approach and interpret the survey content and results. It also provides frame of reference for the survey development team, who should zero-in on this guidance and return to it throughout the item writing process to remain on target and recage whenever temptation for alternative content possibilities arise.

As an example, the PE Surveys opened with:

"The purpose of this research is to solicit information from aircrews and maintainers, regardless of whether or not they have experienced a PE. The results of the survey will be used to focus the investigation into areas that show the greatest promise for determining the root cause(s) of PEs."

Consistent with defining the survey's purpose, it may be beneficial for survey developers to formulate a data analysis plan in advance (or concurrent with) the item writing process. A data analysis plan will clarify the specific study aim(s) each item is intended to address as well as the statistical mechanism(s) by which the item will be analyzed. Use of such a tool will help reduce item creep during survey development, and facilitate analysis on the back end. For additional guidance on data analysis plans, see Panter (2010) and Simpson (2015).

Lesson Learned #4: Item content and structure will influence responses.

There are numerous how-to guides on the "do's and don'ts" of survey design, many of which cover similar ground. To avoid getting into the weeds here, the main lesson learned is that both item content (what is asked) and item structure (how its asked) will influence the manner in which users will respond. Most surveys offer little opportunity for follow-up, and many are completed anonymously; therefore, survey developers must take great care in evaluating each item to ensure it will be interpreted as intended and that it will generate responses that answer the ultimate question.

Initial considerations are whether to use open-ended or closed-ended (e.g., rating scale) items. Open-ended items provide respondents opportunity to comment freely on the question and tend to elicit rich and detailed information; however, they also take more time to complete as well as to analyze (i.e., they create more work for respondents and analysts) and may not always address the question's intent. Closed-ended items, on the other hand, provide a limited number of fixed responses, and can be either dichotomous (e.g., true/false, yes/no) or multiple-choice (e.g., rating scale, check-all-that-apply). While easier to quantify and analyze, closed-ended items provide limited insight beyond their specific subject of inquiry.

For the PE Surveys, our NPS Team took a blended approach of using rating scale items followed by optional open-ended questions. All items were carefully written and reviewed to avoid leading or misleading phrases (e.g., rate your level of comfort instead of how uncomfortable was your…); value-laden terms (e.g., optimal, excessive); absolutes (e.g., never, always), and ambiguous expressions (e.g., often, consistently). We also scrubbed each item for jargon that may not be familiar to all respondents (e.g., most experienced aviators are fluent in aviation terms/phrases, but some students from the T-45 pipeline may still be learning).

For the closed-ended items, we opted to use a 7-point response scale with higher values representing "more" (e.g., higher likelihood, greater frequency). The 7-point scale was selected in order to provide a true midpoint value, with the end points representing all (7) or nothing (1). We therefore assigned descriptive anchors to the 1, 4 and 7 point values (e.g., 1 = not at all likely, 4 = somewhat likely, 7 = highly likely), with no anchors assigned to values 2, 3, 5, or 6, as we determined respondents would be able to discern on their own the difference between 4-5 or 5-6. When relevant, we also included response options for "I do not know" or "Not applicable."

For our data analysis plan, we took two approaches. First, we analyzed all response options independently. Next, we combined ratings of 2-3 and 5-6, thereby artificially deriving metrics for a 5-point scale. Ultimately, this strategy yielded no meaningful benefits. In hindsight, we acknowledge that using a 5-point scale would have been preferable from the beginning. This is consistent with best practices for both unipolar and bipolar scales. For additional guidance on item writing and survey design, see DeVellis (2016) and Maroney & Cameron (2019).

Lesson Learned #5: Tradeoffs are part of the process.

All surveys have their limitations. The NPS PE Surveys, in particular, were very long (~150 items). The Survey Team worked hard to balance the need to limit the length with the need to obtain essential information from respondents. However, we were able to infer from the raw data that a number of respondents "ran out of gas" before making it to the end of the surveys (e.g., stopped responding before completion; limited open-ended responses on latter items). Aside from length, we were additionally mindful of the delicate task to gather sensitive data without inducing stress or calling to mind traumatic events. A review of the open-ended responses suggests our cautionary approach worked in general, though some respondents used language to indicative of emotional provocation (e.g., ALL CAPS, accusatory tone, profanity).

Tradeoffs are inherent to survey design and management. Potential users will be attracted to surveys based on their familiarity, but also because surveys are low cost, capable of quick and widespread dissemination, and relatively easy to analyze. Surveys can also be completed anonymously, which is beneficial in particular when the subject matter is sensitive. However, there are downsides to surveys as well. For instance, their voluntary nature tends to produce a fairly low response rate (typically 20-30%); those who do respond may not fully represent the population of interest (e.g., respondents may hold more extreme views than the majority); and answers can be highly influenced by recent events (e.g., viewpoints on PEs may be more severe in the immediate aftermath of a critical event as opposed to months or years later). Surveys are also subject to systematic error (e.g., a poorly worded item may cause widespread misinterpretation) and random error (e.g., participants responding late at night may misread items due to fatigue).

Final Thoughts

Surveys should be but one of many tools in the Aerospace Experimental Psychologist's toolbox. They are valuable for gathering insights on people's beliefs, opinions, experiences, and perceptions – but should never be considered as a gateway to ground truth. Responsible survey developers should not only be familiar with the strengths and limitations of surveys, but also ensure they appropriately manage stakeholders' expectations on what can and cannot be concluded from survey results. This can be achieved through an open dialogue with the survey sponsors as well as potential end users. Stakeholder participation throughout the survey development process will serve to clarify the purpose, thereby increasing alignment between the items, responses, and the sponsor's intent; provide opportunity for item feedback and revision, resulting in a higher-quality product; and ultimately increase buy-in toward the process, results, and final recommendations. There is an art and a science to the development and use of survey measures, and tradeoffs (e.g., in item content, structure, length, etc.) are part of the process. To help navigate this decision space, it is recommended that future survey developers consult with the community of interest to gather lessons learned and capitalize on shared experience. To what extent do you agree?

References

DeVellis, R. F. (2016). Scale development: Theory and applications. Sage Publications, Inc.

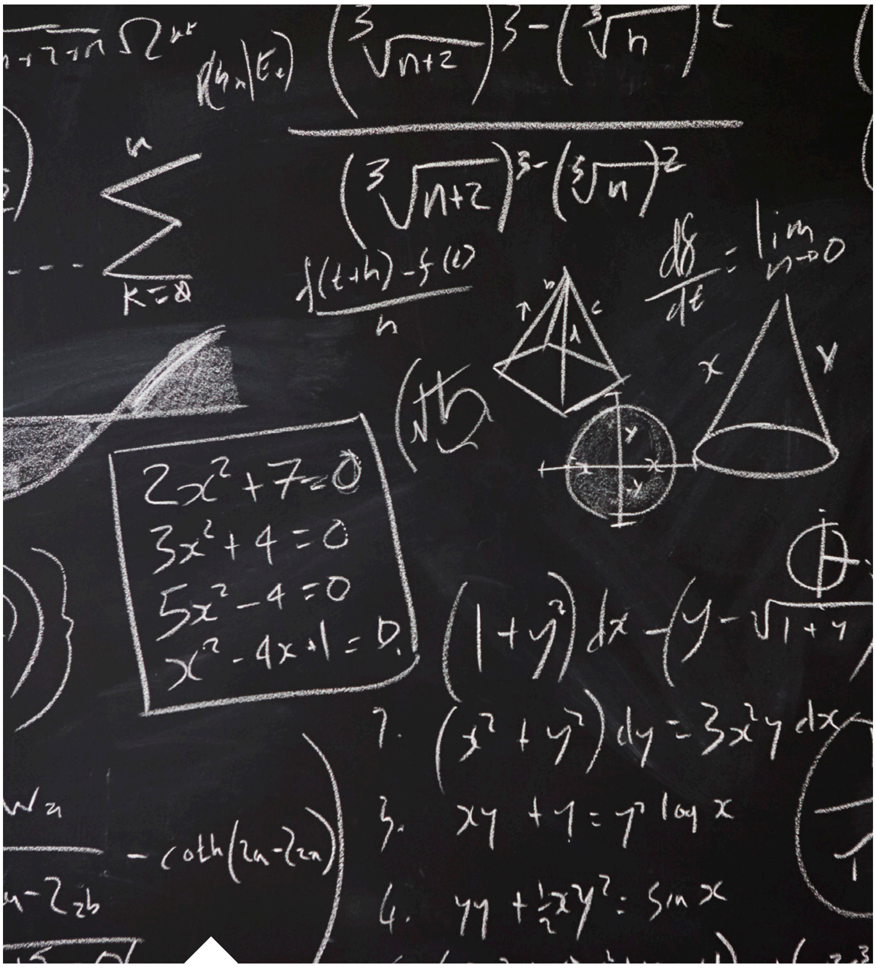
Maroney, W. F., & Cameron, J. A. (2019). Questionnaire design: How to ask the right question of the right people at the right time to get the information you need. Human Factors and Ergonomics Society.

NPS Physiological Episodes Survey Team (2018). Physiological episodes in F/A-18 and T-45 aviators: Report of survey results (Report No. NPS-18-N390-A). Monterey, CA: Naval Postgraduate School.

Panter, A. T. (2010). Writing the data analysis plan. In W. Pequegnat, E. Stover, & C. A. Boyce (Eds.), How to write a successful research grant application (pp. 293-298). Springer.

Simpson, S. H. (2015). Creating a data analysis plan: What to consider when choosing statistics for a study. Canadian Journal of Hospital Pharmacy, 68(4), 3111-317.

QUESTIONS ANSWERED: THE BAYESIAN WAY



Communicating high level concepts and delivering quantitative results can sometimes be an exercise in communication as much as it is an exercise in mathematics. Choosing the right tools that output results in a manner that people can easily comprehend is a challenge. Bayesian statistics are one such method whose outputs are easier to understand by non-statisticians, and can communicate concepts like risk in a more appropriate and meaningful way to better support decision making.

by CDR Jefferson D. Grubb and Andrew N. Koch

AEPs frequently note that the best thing about their job is a strong sense of purpose. What is that purpose? We argue that it is *to answer the Boss's questions*. When analyzing data, AEPs need to use statistics that answer the Boss's questions in a way that the Boss finds compelling. We recommend Bayesian techniques.

Admittedly, most AEPs think their purpose is bigger than simply answering someone's questions. They think it is something like, in ADM Richardson's (2018) words, "[To] protect America from attack, promote American prosperity, and preserve America's strategic influence." AEPs are Sailors, so that answer is accurate but unspecific. All Sailors work to protect America in some way. AEPs do so by applying behavioral science to solve the Fleet's human systems problems. However, AEPs themselves rarely have the authority to implement solutions. Instead, AEPs must convince the Boss, by which we mean some key Naval stakeholder with that authority, to implement the solutions. If an AEP fails to convince the Boss, the Boss won't implement the solutions and thus the AEP won't protect America. Convincing the Boss means answering his or her questions.

For AEPs, answers come from statistics. When people naïvely encounter AEPs, they tend to assume that AEPs are flight-oriented clinical psychologists. Less charitably, AEP community founder CAPT Jack Jenkins (1948) observed that the people he encountered assumed that AEPs were "amateur psychiatrists." However, Jenkins reported that such people were "...much mollified to learn that we were essentially finger counters. And they began to like us better when they realized that our kind of finger counting could often be used to fill out gaps in their own best efforts."

This makes sense. The Boss can get answers from many kinds of experts. These include operational subject matter experts, engineers, financial specialists, etc. What unique insight can AEPs contribute? One might assume that as behavioral scientists, AEPs' unique value lies in their accumulated knowledge of how people think and act. However, operational subject matter experts also have accumulated knowledge of how people think and act. Moreover, they accumulated this knowledge by observing their shipmates thinking and acting in the actual operational environment. If the Boss weighs competing answers that are derived from mere opinions, the opinion informed by operational experience will likely win. AEPs' answers compete because they are derived from formal analysis of relevant data, not merely from opinion. That is, the AEPs' answers are potentially valuable because of statistics.

We say "potentially valuable" because statistics can go wrong in many ways. Most basically, those who don't adhere to the assumptions of their statistical procedures are practicing numerology, not statistics. However, even methodologically sound analyses can fail to convince the Boss. For example, our research team (Beaubien et al., 2015) won a best paper award at I/ITSEC for an analysis that got us laughed out of the room when we briefed it to the Chief of Naval Air Forces. Although scientific reviewers found our study clever and compelling, the CNAF staff thought our data fundamentally couldn't answer the question that the Air Boss had asked.

What was our key failure? Initially, we were tempted to say it was poor experimental design. However, peer reviewers who examined our methods gave us an award. Rather than a poor design per se, our problem was that we didn't adequately consider what kind of evidence the Boss would find compelling. To be successful, AEPs must consider things from the Boss's perspective.

The Boss's perspective comes largely from the nature of the job. Above all else, the Boss makes decisions. Indeed, AEPs need to convince the Boss to implement their America-protecting ideas because the Boss's job essentially consists of deciding what ideas get implemented. Thus, AEPs' analyses need to clearly speak to the decision AEPs want the Boss to make. The price of decision authority is that the Boss is responsible for the cost, schedule, and performance implications of his or her decisions. Failure along any of these dimensions can cost the Boss his or her Boss-hood. AEPs' work usually focuses on performance measures, but their analyses should support an examination of the cost and schedule impacts of the recommended decision. Finally, decisions have deadlines. When a decision is due, the Boss has to make a call regardless of remaining uncertainty. Because the Boss usually lacks time or other resources to collect more data, AEPs should use statistics that permit meaningful interpretation under the widest range of possible outcomes.

To see how different statistical approaches answer the Boss's questions, we conducted a Monte Carlo simulation of a hypothetical training systems experiment. In our hypothetical situation, the mean flight time required for crews to achieve a particular qualification is 10 hours, with a standard deviation of 1 hour. The Boss is considering whether to implement a new training system. Some of the Boss's other experts, hereafter called "bean counters," have calculated that the new training system must save at least 0.5 flight hours on average to be worth the investment. An AEP secures funding, permissions, etc., to measure how long it takes crews who have been provided with a prototype of the new training system to achieve the qualification. Due to time, funding, and other logistical constraints, the AEP is only able to measure 10 such crews before the Boss has to make a decision.

To simulate this scenario, we repeatedly drew samples of 10 values from a Gaussian distribution with a standard deviation of 1.0 hours and a mean set to one of three "effectiveness levels." These levels represented a new training system whose effectiveness was the same (μ =10.0 hours), marginally more effective (μ = 9.75 hours), or meaningfully more effective (μ = 9.25 hours) than the legacy system.

We submitted these data to two kinds of statistical tests. First, we conducted a one-sample Student's t-test, which yielded the 95% confidence interval for the sample. Second, we conducted a naive Bayesian analysis using a t distribution as outlined by Kruschke (2015, pp. 449-472). This yielded a posterior

Table 1. Confidence Intervals and Credible Intervals Across Many Samples

Effectiveness Condition	True Mean	Number of Samples	95% Confidence Intervals Containing True Mean	95% Credible Intervals Containing True Mean
Same as Legacy	10.00	1000	956	961
Marginally More Effective	9.75	1000	938	953
Meaningfully More Effective	9.25	1000	945	958
Total		**3000**	**2839 (94.6%)**	**2872 (95.7%)**

probability distribution from which we computed a 95% credible interval. We repeated this procedure 1000 times for each of the three effectiveness levels.

Before examining our results, we must consider the subtle differences between these two kinds of statistics. Frequentist statistics, such as Student's t, define probability as the frequency of a particular result given many repeated samples. Thus, a 95% confidence interval is an interval that is calculated according to a procedure that, if calculated on many samples, would contain the true mean 95% of the time (Neyman, 1937). Importantly, this is different from saying that there is a 95% probability that any single interval contains the true mean. Jerzy Neyman (1937), who invented confidence intervals, specifically warned against making such an interpretation. Instead, a confidence interval is essentially an inversion of a p-value. It identifies the span of null values the researcher would have failed to reject.

In contrast, Bayesian statistics define probability as the degree of confidence or belief in a particular result given prior beliefs and the observed data. Thus, a 95% credible interval is the shortest interval of the posterior distribution that contains 95% of the probability mass. This means exactly what it sounds like it means: the statistician is 95% confident that the true mean lies in the interval. However, statisticians continue to debate what it means to be "95% confident" in different contexts. Also, researchers will calculate different posterior distributions if they use different priors. In practice, Bayesian researchers invoke flat, or "uninformed," priors to mitigate this criticism, but doing so reduces the effective power of the statistic. We used uninformed priors for our Bayesian analyses.

To see how well our statistics delivered what they claimed, we counted the number of samples in which the 95% confidence intervals and 95% credible intervals contained the true mean. As may be seen in Table 1, the 95% confidence intervals contained the true mean on 94.6% of the 3000 total samples. Similarly, the Bayesian credible intervals contained the true mean on 95.7% of the same samples. Thus, over 3000 samples, both the t-test and Bayesian analysis delivered results within 1% of the predictions of traditional (i.e. frequentist) interpretation of probability, presumably allowing practitioners of both statistical philosophies to have great confidence (in the Bayesian sense) in their results.

However, the real question is what these statistics can tell AEPs, and thus what AEPs can tell the Boss, on any one sample. To find out, we counted the number of samples on which each kind of analysis supported the correct decision. For the "meaningfully more effective" condition, this meant that the intervals were entirely below 9.5. For the other two conditions, this meant that the intervals were entirely greater than 9.5. If an interval from the "meaningfully more effective" condition was entirely greater than 9.5 or an interval from the other conditions was entirely less than 9.5, we counted it as incorrect. We deemed intervals that contained 9.5 to be inconclusive. As may be seen in Table 2, both kinds of 95% intervals proved fairly conservative. They both supported definitively wrong decisions less than 1% of the time. However, they also supported definitively correct decisions infrequently. Most of the time, they would leave the AEP strictly without much to say.

This shouldn't be surprising. An *n* of 10 is fairly small, though not uncommon for many projects that AEPs work. Still, when AEPs say, "The results were inconclusive," the Boss is prone to hear, "I spent your money and wasted your time." If AEPs used frequentist statistics, they have little recourse. The proximity of the critical value to one end of the interval makes no difference. AEPs could compute a more liberal confidence interval (e.g. 75% instead of 95%), but only if they decided to do this prior to any analysis. Remember that the percentage of a confidence interval refers to the performance of a particular calculation over many samples. Recalculating intervals until you find one that excludes the critical value on any given sample is an entirely different procedure that has no grounding in any definition of probability. Those who do this are numerologists, not statistician.

Table 2. Decision Accuracy of 95% Confidence and 95% Credible Intervals

Condition	95% Confidence Interval *Right*	*Wrong*	*Inconclusive*	95% Credible Interval *Right*	*Wrong*	*Incolculsive*
Same as Legacy	314	0	686	278	0	722
Marginally More Effective	111	2	887	95	1	904
Meaningfully More Effective	110	6	884	79	4	917

In contrast, Bayesian analysis leaves the AEP with valid routes to a partial answer. The posterior distribution that underlies a 95% credible interval is a probability distribution. If the critical value lies inside of the credible interval, the AEP can calculate how much of the probability of the interval lies on either side of it. For that matter, the AEP can validly report what percent of the entire probability distribution is on one side or another. In our scenario, even if the credible interval contained 9.5, a Bayesian AEP would still be able to tell the Boss something like, "It's 75% probable that the new system will save at least 0.5 flight hours." Equivalently, the AEP could validly report, "The odds that the new system is worth buying are 3 to 1." The Navy values calculated risk taking and so the Boss speaks this language, probably natively. He or she might not like those odds, but they give him or her a better basis for a decision than, "I don't know."

In conclusion, both frequentist and Bayesian statistics can inform decisions, even when those decisions are partly dependent on factors other than the data that are directly analyzed. However, Bayesian techniques permit a researcher to validly say something meaningful under a wider array of possible experimental outcomes. In this paper, we are not arguing that Bayesian techniques answer the Boss's question better or worse than frequentist techniques. However, we do argue that Bayesian techniques answer the Boss's question more often. If AEPs want to protect America, this is an important consideration.

Beaubien, J.M., Stacy, E.W., Wiggins, S.L., Keeney, M.J., Bolton, A.E., Grubb, J.D., Walwanis, M.M., Priest, H., & Riddle, C.S. (2015, November). Differentiating measures of learning from measures of performance during aircraft carrier landing practice. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Jenkins, J. G. (1948). The nominating technique as a method of evaluating air group morale. The Journal of Aviation Medicine, 19(1), 12-19.

Kruschke, J. K. (2015). Doing Bayesian Data Analysis, a Tutorial with R, JAGS, and Stan (2nd ed.). Academic Press. https://doi.org/10.1016/c2012-0-00477-2

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 236(767), 333-80. https://doi.org/10.1098/rsta.1937.0005

Richardson, J.M. (2018). A design for maintaining maritime superiority 2.0. Office of the Chief of Naval Operations. https://media.defense.gov/2020/May/18/2002301999/-1/-1/1/DESIGN_2.0.PDF (Accessed 30 September 2020).

CALL SIGNS is a bi-annual publication of the US Navy Aerospace Experimental Psychology Society (USNAEPS).

Staff:

LT TODD SEECH/ NEWSLETTER EDITOR IN CHIEF.

LT CLAIRE MODICA/ ASSISTANT EDITOR.

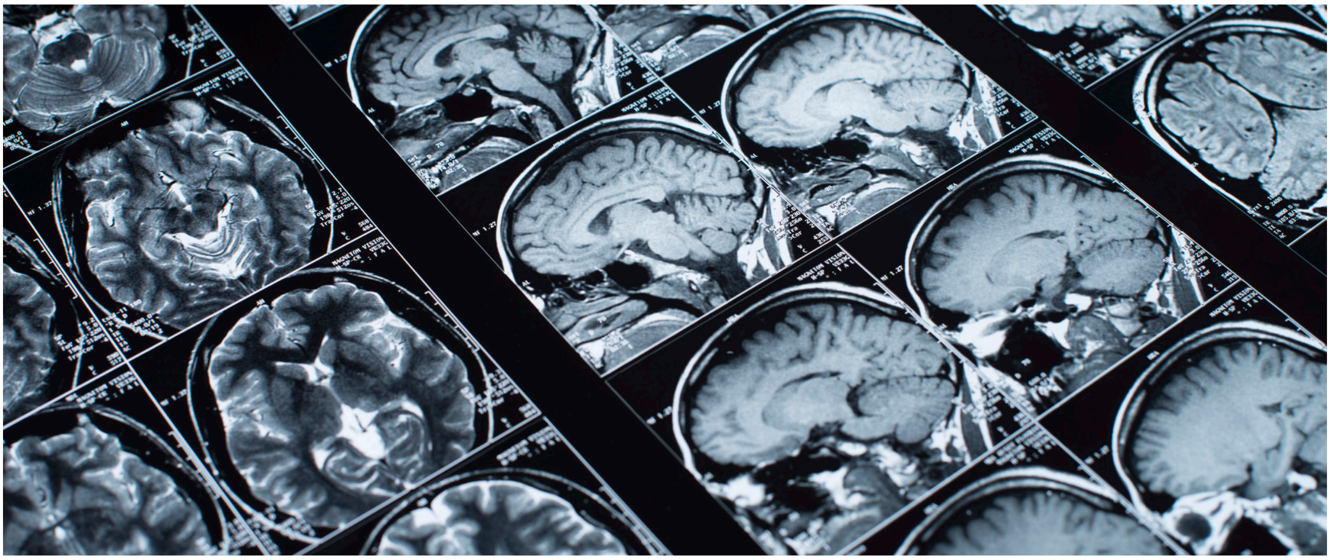
LT ERIC S. VORM / LAYOUT EDITOR AND GRAPHIC ARTIST.

U.S. Navy
Aerospace Experimental Psychology

TAKE YOUR CAREER TO NEW HEIGHTS

The US Navy Aerospace Experimental Psychology community is seeking qualified applicants for positions in research, development, testing and evaluation. Selected candidates will receive a commission in the United States Navy as Lieutenants (O-3), and will attend Navy flight school. Qualified applicants must hold a doctorate in Human Factors Engineering, Industrial/Organizational Psychology, Cognitive Psychology, Neuroscience, or related fields.

Apply your skills and training in the dynamic, fast-paced, and exciting field of Aerospace Experimental Psychology. **Apply today!**



THE VALUE OF RESEARCH & DEVELOPMENT

by CAPT Joseph V. Cohn

The decision to invest resources in a new research and development effort is complex. Decision makers, whether they are in government, academia, or industry, are obliged to ensure that these investments have value, ideally in advance of committing significant resources. This is particularly important in sectors in which such efforts are high risk and expensive and require lengthy processes that significantly extend the time required to transform outcomes into marketable capabilities. The biomedical sector is an example for which it is critical to develop methods and techniques for determining at the start of an effort the return on future research and development investments. In this sector, the cost of transforming research and development outcomes into tangible products, like biologicals and pharmaceuticals, continues to rise, while the time available to effect this transformation continues to decrease (Grazier et al. 2005). Knowing in advance where best to place investments would help reduce the complexity of these decisions. The most common approach to calculating the value of an investment is through a Return on Investment calculation (RoI; Cohn & Mutai, 2019). Yet, RoI is simply a measure, like temperature or weight, and does not indicate a research investment's intrinsic 'value' (Cohn & Fletcher, 2010). Value varies based on the context within which it is being assessed. For an industry seeking to recoup their initial investments and secure a profit, value reflects the total revenue anticipated for a product minus all estimated expenses needed to market the approved product. For other entities, like governments, value may reflect more esoteric outcomes, like gains in overall knowledge, increased employment opportunities, and quality of life improvements (Health Economics Research Group, 2008). In this article, we explore three different frameworks that value research investments in different ways and provide examples to demonstrate their utility. These frameworks flow from (1) a basic RoI calculation which provides minimal information but in an easy to calculate manner, to (2) an RoI calculation that is more aligned to the manner in which research investments are actually made – annually, from different fiscal sources, and in parallel to other related efforts, to (3) an RoI calculation that embeds cost benefit tradeoffs. Collectively these three frameworks highlight the importance of clearly articulating the context within which the RoI "question" is posed, in order to determine the best approach to use and to interpret the resultant outcomes.

RETURN ON INVESTMENT FRAMEWORKS

Basic RoI Calculation

Phillips (2003) and others propose calculating RoI as a simple ratio between the net value of the investment (i.e., after the cost of the investment is subtracted out) divided by the cost of the investment, yielding net return for each unit of investment (Figure 1). RoI is usually calculated for some period of time, such as a year and can be represented either as a ratio or as a percentage. Figure 1 shows RoI calculated as a ratio. Multiplying by 100 would convert this value to a percentage.

$$RoI = \frac{\text{(Value of Investment) - (Cost of Investment)}}{\text{Cost of Investment}}$$

Figure 1: Basic return on investment calculation (after Phillips, 2003; Cohn & Fletcher, 2010)

As one example, consider how this calculation may be used to confirm the value of investing in developing a new vaccine to combat a virus that, unchecked, leads to significant loss of duty days (after Berst & Bennet, 2017; see also Radin et al 2014). We assume the following:

·The product was proven to work, and approved for use, in FY 2012 (it is now FY 2020). The time period over which this product is used is therefore 8 years.

·Total research investment prior to FY 12 was $107M. Note that this investment would also include Development and Regulatory costs.

·The cost of lost duty days, which can be calculated using Military Composite Standard Pay and Reimbursement Rates based on known lost duty days, is on average $30M per year.

The calculation is then:
This yields an RoI of 1.24. In other words, this research investment returns one dollar and 24 cents for every dollar invested.

$$RoI = [(\$30m \; x \; 8) - \$107m] \, / \, 107m$$

Note that in this example this approach to calculating RoI doesn't indicate the true value of the research investment in terms of a need, a new capability, and a replacement capability. One could propose that preventing a sufficiently large number of lost duty days would be significant, so that a ratio larger than 1.00 would be satisfactory. Yet, absent additional information it is not possible to conclude what the specific RoI would need to be in order to justify the investment.

Net Present Value Calculation

Net Present Value (NPV) is a method of calculating the value of an investment made over time. NPV is ideally suited to an effort focused on delivering a new capability that will be phased-in to replace an existing capability because it is anticipated to be more effective. As one example, consider a research effort to deliver a new capability, termed a Digital Tutor, to provide training that reduces the time to transform novices into experts from seven years to one year, using a phased-in approach (Cohn & Fletcher, 2010).

In order to effectively calculate NPV, a proper accounting of the costs must be made. Here, this accounting focuses on comparing the costs for the traditional training system compared to the costs for the new training system, taking into account the phased-in approach. This must be done by first clarifying the total investment horizon – the total number of years over which the calculation will be made. Here this horizon includes time needed to completely phase-in the new capability. In this example, the horizon was determined to be twelve years. The first four years capture the research and develop investments. An additional year is given to support ramp up of the new capability and being integrating it into instructional use. The remaining seven years are included because the working hypothesis is that the training will produce in one year students who are performing at the level of individuals with seven years of experience. The value of this new capability therefore is the reduced cost of on the job training (OJT) by more quickly developing experts.

Figure 2 shows the two different investment streams that are being compared through the NPV approach. The top row, "A School" depicts the traditional training approach for each of the 12 years being considered. Every year a new group of students receive "A" school training and then continue to receive on the job training (OJT) to achieve the level of expertise expected of students receiving only Digital Tutor instruction. The bottom row, "A School Plus Digital Tutor" shows the new training approach. Here, for the first four years students receive "A" school and then continue to receive OJT. At the same time research funds are invested in developing and testing the Digital Tutor. In the fifth year, the Digital Tutor is implemented, and "A" school is no longer provided. The OJT costs of students who had received "A" school training in the first four years prior to the Digital tutor continue to be accounted for, but these costs taper each year as more students who have received the Digital Tutor training enter the workplace. This tapering is noted in years 6-11 as the OJT values decreases from "6" to "1" – the last student receiving only "A" school instruction begins their seven years of OJT in year 5, and for tracking purposes would complete that OJT at the end of Year 11.

When all associated costs are for each investment stream are calculated, the NPV difference – the benefit of using the new training capability compared to the traditional capability – is $1.3B. Compared to the Basic RoI Calculation, the NPV approach allows for a more detailed inclusion of different types of costs, and requires more rigor from the outset. Using this approach requires laying out the factors contributing to the costs of the proposed investment, and setting a correspondingly detailed baseline against which to compare them. Yet, as with the Basic RoI Calculation, the resultant value is just that – a value. Absent additional information for comparison, it is not possible to conclude if this RoI is of great value, minimal value or no value.

Training Options	YEAR											
	1	2	3	4	5	6	7	8	9	10	11	12
"A" School	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7	A OJT/7
"A" School plus Digital Tutor	A OJT/7 DTD	A OJT/7 DTD	A OJT/7 DTD	A OJT/7 DTD	– OJT/7 DTD	– OJT/6 DTO	– OJT/5 DTO	– OJT/4 DTO	– OJT/3 DTO	– OJT/2 DTO	– OJT/1 DTO	– – DTO

Figure 2: **A** = "A" School yearly operating costs | **OJT/n** = Yearly on-the-job training costs for n*2000 students | **DTD** = Yearly DT development costs | **DTI** = DT start up and outfitting costs | **DTO** = DT yearly operating costs

COST BENEFIT ANALYSIS

A different approach to valuing research investments is to ask whether or not the benefits are worth the cost. This perspective extends the results of both previous examples, reframing the RoI question in terms of "if this solution works, how long would it take to recoup the investment?" Or, posed differently, it allows us to ask when the breakeven point, the point in time at which the benefit equals the cost, is expected to occur. As with the NPV approach, this approach requires us to make assumptions about the effectiveness of the proposed investment.

Consider the first example. Here the breakeven point would tell us how many years of sustained reductions in lost duty days are needed to justify the research investment. The breakeven point, BP, can be calculated as (Figure 3):

$$BP = \text{Res Cost} / (\text{Loss Cost X Reduced Risk})$$

Figure 3: Breakeven point (BP) calculation. **Res Cost:** Total research investment cost **Loss cost:** Cost of lost duty days | **Reduced Risk:** Reduced in risk (here, effectiveness of the treatment in terms of reducing the number of duty days lost to infection). This is expressed as 1 - ($Risk_e$ / $Risk_0$) where $Risk_0$ is the observed risk and $Risk_e$ is the expected risk. Adapted from Mueller & Stewart, 2011.

Reduced Risk is the variable allows us to value multiple hypothesis regarding the effectiveness of a given approach. Here, we know RiskO to be 50,000 lost days per year. RiskE can be varied to allow us to test multiple thresholds for our treatment. Consider setting RiskE to 200 days – a bold statement that our intervention would almost effectively reduce the number of lost duty days to 0. In this case, BP turns out to be 3.6 years, which means that for an investment of $107M, we can essentially eradicate the consequences of infection and recoup our investments in 3.6 years.

The question for a decision maker would be whether or not they could absorb the investment cost over the stated time period.

CONCLUSION

Increased time to transform research into commercial products and competing priorities combine to place a unique burden on decision makers, thereby requiring them to make investment decisions with incomplete information. Return on Investment calculations, like those presented here, can help reduce the complexity of these decisions by providing different ways of valuing these investments. While no single approach, or set of approaches, will provide a "crystal ball" to enable decision makers to decide with perfect certainty, the approaches presented here do offer additional quantifiable methods to reduce some of complexity surrounding investing in research.

REFERENCES

Cohn, J. V., and Fletcher, D. (2010). Interservice/Industry Training, Simulation and Education Conference, At Orlando FL, Volume: Proceedings of the 31st Interservice/Industry Training, Simulation and Education Conference. Orlando FL.

Cohn, J. V. & Mutai, J. C. (2019). Introduction to the thematic issue on valuing research: theory, research, policy and practice of return on investment. Theoretical Issues in Ergonomic Sciences.

Grazier, K. L., Trochim, W. M., Dilts, D. M., & Kirk, R (2013). "Estimating Return on Investment in Translational Research." Evaluation & the Health Professions 36(4): 478-491.

"Health Economics Research Group, Office of Health Economics, RAND Europe. Medical Research: What's It Worth? Estimating the Economic Benefits from Medical Research in the UK." RAND Corporation, 2008. https://www.rand.org/pubs/external_publications/EP20080010.html.

Mueller, J. & Stewart, M.G. (2011). Balancing the Risks, Benefits, and Costs of Homeland Security. Homeland Security Affairs, https://www.hsaj.org/articles/43.

Phillips, J.J. (2003). Return on Investment in Training and Performance Improvement Programs, (2nd Edition). Oxford, UK: Butterworth-Heinemann.

Radin, J. M., Hawksworth, A. W., Blair, P. J., Faix, D. J., Raman R., Russell, K. L. and Gray G. C. (2014). Dramatic Decline of Respiratory Illness Among US Military Recruits After the Renewed Use of Adenovirus Vaccines. Clin Infect Dis. 59(7): 962-968.

ANNUAL USNAEPS AWARDS

MICHAEL G. LILLIENTHAL LEADERSHIP AWARD

The CAPT Michael G. Lilienthal Award for Leadership is awarded in recognition of an individual who has significantly advanced the field of Aerospace Experimental Psychology through excellence in leadership over the past year. Award recipients have consistently demonstrated their ability to: motivate and inspire others; apply foresight and resourcefulness in anticipating and overcoming significant challenges; maintain strength of character in the face of adversity.

LCDR Brennan D. Cox is recognized for outstanding leadership as Deputy Director, Naval Aerospace Medical Research Laboratory, Naval Medical Research Unit Dayton. LCDR Cox exercises delegated cognizance over the mission of NAMRL, the US Navy's flagship aeromedical research laboratory, guiding the work of 75 scientists, engineers, and support staff in executing the laboratory's scientific mission. LCDR Cox

demonstrated outstanding leadership as NAMRL Deputy Director, expertly leading the development of strategic initiatives, forging key partnerships with new research collaborators and stakeholders, and overseeing scientific products generated by staff scientists, with NAMRL achieving its most productive year since relocating to Dayton.

ROBERT S. KENNEDY EXCELLENCE IN AVIATION RESEARCH AWARD

The CDR Robert S. Kennedy Award for Excellence in Aviation Research is awarded in recognition of an individual who has made significant and outstanding contributions to the field of aerospace psychology through original research over the past year. Award recipients have consistently demonstrated their ability to apply their scientific acumen to solving research challenges of critical importance to the Naval Aviation community. The results of their research have directly contributed to demonstrably more effective Selection, Training, Safety and Human Performance technologies in the service of Naval Aviation.

LT Sarah M. Sherwood is recognized for the Robert S. Kennedy Award for Excellence in Aviation Research in recognition of numerous and impactful research contributions in the areas of aviator safety, health, and training performance. During this period, her test and evaluation work led the U.S. Coast Guard to procure innovative laser eye protection for all their aviation platforms to combat emerging airspace threats. She also played a critical role investigating counter-

measures to motion sickness that will inform joint aeromedical acquisition decisions to avoid lost training time and attrition for reasons of aeronautical adaptability. Furthermore, her research on spatial disorientation training tools will revolutionize how aviators build awareness of the leading causal factor of catastrophic flight mishaps.

PAUL CHATELIER LIFETIME ACHIEVEMENT AWARD

The CAPT Paul R. Chatelier Award for Lifetime Achievement honors individuals who have significantly and uniquely shaped the field of Aerospace Experimental Psychology through scientific, analytic, managerial and leadership excellence over the course of their career. Award recipients have demonstrated a broadness of vision combined with force of character to achieve long ranging goals that have often run counter to common wisdom. The results of their dedication, persistence and foresight have led to

paradigm shifting accomplishments that enable the Naval Aviation community to rapidly and effectively overcome current and emerging challenges and threats.

CAPT (ret) Russell Shilling is recognized for significant and enduring contributions in the areas of aviation human factors, Science, Technology, Engineering, and Math (STEM) education, and psychological health. He successfully leveraged industry standard technologies such as immersive virtual reality and artificial intelligence to develop innovative military training simulations and PTSD treatments that helped launch the "games for impact" movement. In collaboration with Sesame Street Workshop, he developed award-winning programs to help children of service members deal with stress, grief, and deployments. A tireless advocate for STEM education, he served as the U.S. Department of Education Executive Director for STEM initiatives and a representative for the National Science and Technology Council (NSTC) Committee on STEM.

Call Signs is an official publication of the US Naval Aerospace Experimental Psychology Society, and is published twice a year.

FAIR WINDS AND FOLLOWING SEAS

CDR Hank Phillips, AEP #119

CDR Hank Phillips joined the Navy in 2001, and was winged as AEP #119 on January 25th, 2002. He served in a variety of duty stations and roles, including serving as AEP specialty leader from 2019 - 2020. We sat down with him on the eve of his retirement to discuss what he has learned from 20+ years in the Navy, and take a look at his remarkable career.

What made you want to become an AEP?

To be honest, I didn't even consider a career in the military until I was almost done with grad school. At SIOP in 2000, I met a researcher named Paul Bliese, who was doing some really neat stuff in R at the time, and was an Army Major. I started conversations with him about a possible career in the Army, and while that process was still underway, I remarked to my wife that while this possible career as a scientist in uniform sounded exciting, it was a shame that there was no way to do this and fly airplanes. She did a search on monster.com that night and found an ad for the AEP community, which was looking for I/O PhDs who were willing to learn to fly. I called [Assistant Specialty Leader] LCDR Sean Biggerstaff the next day, and a month later, I was in the Navy. I had no idea at the time that a job like this even existed. I'm proud to say that I think we've done a great job since then of raising awareness about the AEP community among grad student populations in the disciplines we draw from through consistent, intentional presence at conferences and through their recruitment centers.

That said, even at that point I did not expect to take to a career in uniform as readily as I did. The experiences I was able to have in that first tour, and the impact of our work on the fleet, got me hooked for the long term. Life as an AEP was something I tried on, and that grew on me faster than I ever expected.

Looking back to your early career, were there any moments that stood out as shaping your career path?

When I came in, I'd focused almost exclusively in graduate school and my internship on personnel selection work. My loftiest aspiration at the time was to overhaul the Navy's aviation personnel selection process. I was in the right place at the right time, wrote a proposal to do just that as my student capstone project, and was lucky enough to get the effort funded. In my first tour we were also able to get the Aviation Selection Test Battery (ASTB) converted to a theta-metric and develop three new test forms to remedy a large scale compromise, and develop a computer-adaptive engine for multiple-choice ASTB content that's still in use today.

In order to get to that point, I had some remarkable experiences. In my first year,

Hank and his wife, Mimi, Celebrating the Navy's Birthday in 2002.

Hank Phillips sings kareoke in Seattle after the 2019 Human Factors and Ergonomics conference.

I had the opportunity to brief the Chief of Naval Air Training and his Commodores about the ASTB revision effort with my new Department Head CAPT John Schmidt, who'd been in his position for a week at that point. I have told the story of this briefing many times; I learned a lot that day about the differences between academic and military audiences.

By the time I had gotten to my second duty station, which happened to be CNATRA, I knew I was in for the long haul. Duty stations like those offer moments when as AEPs, we are able to make changes to policy or improvements to training that have an immediate impact on the fleet. Those really stay with you.

Did you have any mentors who helped shape your leadership style? How did they influence you?

I had so many mentors, both within and beyond the AEP community, that no matter what I offer here I'll do a disservice to many people who taught me what I know. But I would like to single out three AEPs in particular:

CAPT John Schmidt was the second boss I had in uniform. He had an amazing grasp of the systems that made the Navy run, and was completely unafraid to stand up to right a wrong. He got a lot done in a very short period; his pace is something I emulate to this day.

CAPT Dylan Schmorrow I knew peripherally from my first tour, but really got to know him when I served as his ASL from 2011-2013. That established a mentoring relationship that persists to this day. From him I learned the importance of networks, the ability to see all the pieces on the board, and the ability to see situations from multiple perspectives to understand why people make the decisions they do. He understands people, and is also one of the hardest working people I know.

CAPT Joseph Cohn, whom I've known almost my entire career, is one of the finest strategic planners I've known. His ability to find the sources and information needed to inform big decisions, and to build the teams needed to shape and implement them, is second to none.

These leaders also live the values of integrity, honesty, and courage, like all the finest leaders I've learned from.

Over your career, how has the AEP community changed and where do you see opportunity for the future of the AEP community?

Over the 20 years I was in, the community's skillsets have broadened considerably, as the skills needed to stay on the cutting edge in behavioral science and systems design have evolved. New AEPs frequently know as much about technology and coding as they do about research and methods. It is truly a different world. This skill diversification must continue, in order to keep AEPs in the right rooms and right discussions about where Naval aviation and Navy medicine are going. A broader focus on systems of systems is increasingly important.

Do you have any advice for young officers looking to make a career of the Navy?

I built my remarks at my retirement ceremony around three pieces of advice for young AEPs, to which I've added a couple additional points below:

1. Don't take yourself too seriously, and don't believe your own BS. You are smart and capable, or you wouldn't be in this community. Don't ever make the mistake of thinking that this makes your opinion the only one that matters, or even worse, let yourself believe you understand a fleet problem better than the warfighters who are living it.

2. Be ready to make hard choices, and don't expect to be thanked for them. The right choice is not often the most popular choice. One of the most important responsibilities of a leader is doing what is necessary, not what those around you want to see happen.

3. If you want to succeed in this Navy, you had better get comfortable with relentless self-promotion. Your career depends every bit as much on your ability to produce a record of your accomplish-

Hank Phillips and fellow AEP LT Rick Arnold in Puerto Rico, circa 2004.

ments as it does on making the world a better place. It's your responsibility to do your best, but all those accomplishments will not help you get promoted unless they are reflected in your record. The Navy is no place to be humble.

4. Guard your reputation. You only have one, and when it's gone, it's extremely difficult to get it back. Your promotion success may depend on your officer record, but every other aspect of your career, in uniform and beyond, depends on who trusts you and who takes you seriously. Don't ever lose sight of this.

5. Deploy if you can. And if you can't, find a reason to visit, travel, or serve OCONUS, even if it's only for an applied project. In the post-2020 Navy, the single best thing you can do to help your odds at the Selection Board may be to build a record that makes it clear that you serve where the fleet is. I don't believe it has ever been as important as it is today.

What are your plans for the future?

I started terminal leave in the middle of the pandemic, so I didn't take any time off; there was nowhere to go. I've joined the team at SoarTech as Director of Learning Technologies, working from their Orlando office, with my old friend Dylan Schmorrow. It's a great fit for me, and lets me leverage all that I learned about Team Orlando and joint training and program needs after 5 years there, and all the great relationships and friends I was fortunate enough to make while assigned to NAWCTSD.

"Fair winds and following seas" is the traditional Navy send off for those who are retiring. It reprepresents the crew's best wishes for safe travels and peaceful conditions after a long and sometimes arduous career in the sea service.

Hank Phillips served 20 years as an AEP, serving as both Assistant Specialty and Specialty Leader of the community, as well as numerous other leadership roles throughout the Naval Enterprise.

We say goodbye to one of our own, and remember their contribution to Naval Aviation and the Naval Aerospace Experimental Psychologist community

WE REMEMBER

by Dr. Angus Rupert

Naval Aerospace Experimental Psychologists are scientists in uniform. We perform a unique role in the US Navy--serving as both experts and consultants in human performance, and also researchers. We seek both to contribute to Naval Aviation's current needs, while also providing insights and pathways to future capabilities.

Beginning in the early 1940s, the role of disciplines such as human factors engineering, industrial/organizational psychology, and cognitive neuroscience began to grow in response to a number of challenges forced by the emergence of new technologies such as fighter/bomber aircraft and aerial gunnery. Throughout the decades that followed WWII, AEPs and their skills were applied in increasing areas, including early space exploration. Dr. Manning Correia (AEP #16) was one of these individuals whose expertise proved to be the right solution at the right time. His unique insights and tireless work ethic helped the US Navy in countless ways, and his research--which continues to be cited today--quickly became a standard for excellence in applied military psychology in the engineering sense.

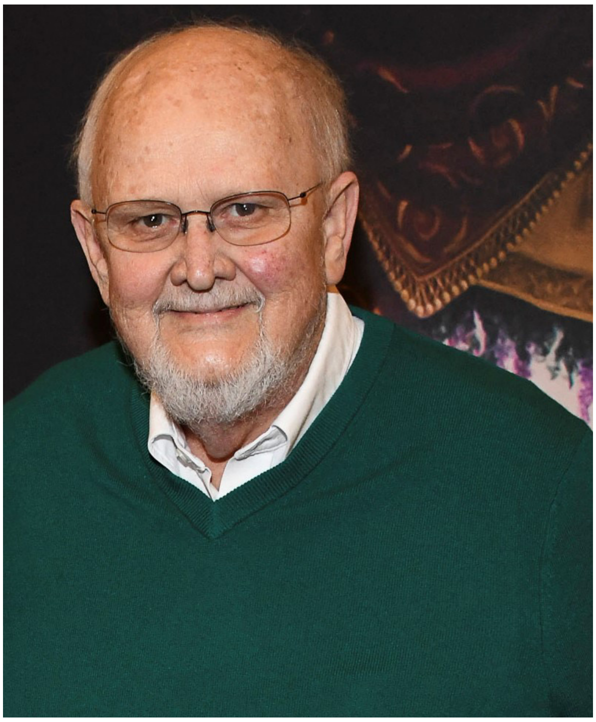
Dr. Manning Correia passed away on November 5, 2020 in Woodlands, Texas. Following his PhD at University of Alabama, Manning joined the U.S. Navy and was assigned to Pensacola during the 1960s when NASA heavily funded the early space program efforts at the Naval Aerospace Medical Research Laboratory (NAMRL). As a fledging Lieutenant, Manning published what is one of the most cited articles from NAMRL –" Hixson, W.C., Niven, J.I., Correia, M.J.: Kinematics nomenclature for psychological accelerations. Monograph 14. Pensacola, Fla.: Naval Aerospace Medical Institute 1966. Vestibular scientists dealing with orientation of the vestibular end organ still reference this classic article.

Manning left the Navy to work and study in Canada at the Defence and Civil Institute of Environmental Medicine (DCIEM) which was a well-known vestibular research center in the early 1970s. On his return to the U.S., he joined the faculty at University of Texas Medical Branch in Galveston where he set up a lab conducting basic neurophysiology of the vestibular system. Throughout his career he maintained connectivity with the Navy and NASA while funded by the National Institutes of Health (NIH).

During the height of the Cold War, despite high tensions between Russia and the US, there was one program where scientists from both countries worked closely together with frequent travel between US and Russia—a rhesus monkey space program with several launches of biological species from Russia. From the mid-1980s to the early 1990s, Manning was the principal investigator of a large scientific team which prepared the primates and designed experiments to record vestibular neurons while in space.

Manning's post-doctoral and graduate students have gone on to star in vestibular research at prestigious universities. Dr. Correia was well funded throughout his career by NIH and when he decided to delve deep into the molecular basis for the perceptual and neurophysiological research he had been conducting for 25 years, the NIH awarded him a rare form of funding reserved for outstanding scientists to change career directions which permitted Dr Correia to move into the biochemical and genetic areas of vestibular hair cell research.

Not all members of the AEP community serve until retirement from the Navy. Many go on to careers in different fields. Manning serves as an example of an AEP who launched his outstanding academic career from his research tour in Pensacola.



AEROSPACE EXPERIMENTAL PSYCHOLOGY

WHERE SCIENCE AND SERVICE UNITE

TAKE YOUR PHD TO NEW HEIGHTS IN THE US NAVY AS AN AEROSPACE EXPERIMENTAL PSYCHOLOGIST!

1- HUMAN FACTORS

We conduct research and analyses of naval aviation platforms and systems, using cutting edge human performance techniques.

2- TRAINING & SIMULATION

We evaluate and develop training technologies for all elements of aviation platforms that feature state-of-the-art in virtual and augmented reality.

3- PERSONNEL SELECTION

We oversee the selection test battery used for identifying ideal candidates for naval aviation by predicting success in the training pipeline.

US naval aviation is a fast-paced and dynamic domain in which you can immediately apply your skills and education while serving your country.

The opportunities and benefits- from generous pay and allowances to once-in-a-lifetime experiences- are abundant.

Come see how you can make a difference! Contact me to learn about our recruitment process at:

www.navyaep.com

Captain Joseph Cohn, PhD
AEP Specialty Leader