EMBRY-RIDDLE Aeronautical University SCHOLARLY COMMONS

International Journal of Aviation, Aeronautics, and Aerospace

Volume 6 | Issue 3 Article 4

2019

The application of Signal Detection Theory principles to aircraft certification

John M. Maris Ph.D
Embry-Riddle Aeronautical University, jmarispq+erau@gmail.com
Alexander V. Ilyin Ph.D
State Research Institute of Aviation Systems, Moscow, Russia, ip.alexander.ilyin@gmail.com

Introduction

The certification of aircraft systems has traditionally been founded on a risk-based approach that balances the severity of the hazards caused by a system failure against the probability of their occurrence. While this approach works well for essential aircraft functions, such as airspeed or altitude indications, it can hamper the adoption of new safety systems. This is because the benefits of such systems are not generally accounted for in the certification process, which is only concerned with the potential hazards and failure probabilities arising from their incorporation.

This paper proposes the application of Signal Detection Theory (SDT) concepts to optimize the risk/benefit ratio for the certification of optional equipment that is intended to enhance aviation safety and/or operational effectiveness. In many cases, the proposed method would lower the certification barriers for the deployment of such systems, leading to potentially significant aviation safety benefits, as exemplified by the introduction of airbags into automobiles.

Air bags were available as optional equipment for passenger cars beginning in the early 1970s, but their installation remained optional until the passage of the Intermodal Surface Transportation Efficiency Act of 1991, which made them mandatory for the front seat occupants of all passenger automobiles and light trucks. Although it was quickly recognized that airbags could convey significant safety benefits, they also carried two major risks: unwarranted deployment, and serious injury or death to vehicle occupants of small stature, such as children. The decision to mandate airbag fitment resulted from their overwhelming benefits, despite these potential drawbacks. Unfortunately, aircraft certification regulations do not use the risk-benefit analysis that led to the widespread adoption of life-saving air bags in automobiles.

Federal Aviation Administration (FAA) Advisory Circular (AC) 23.1309-E provides guidance for the system safety analysis and assessment for Part 23 airplanes. The decision tree incorporated in the guidance addresses adverse effects, failures, malfunctions and hazards, but makes no mention of benefits (FAA, 2011, p. 17). Similarly, Figure 2 of the FAA document defines the “relationship among airplane classes, probabilities, severity of failure conditions, and software and complex hardware and Design Assurance Level” (FAA, 2011, p. 23), but no mention is made of the potential benefits, or their likelihood, of the system being installed. This same risk-based philosophy is carried over to other important advisory material, including the guidance for software certification (RTCA, 2012), complex hardware (RTCA, 2000) and system safety analysis (SAE, 2010).

The following material develops the mathematical basis for the application of SDT and Bayesian methods to the certification of optional aircraft systems. The discussion begins with a review of SDT principles, which are then mapped to their counterparts in the certification domain. The concepts of cost and efficiency are then applied to optimize the risk/benefit ratio for the system under investigation. The discussion concludes with a case study of the method's application to an Electronic Flight Bag (EFB) software application.

Signal Detection Theory Basics

Signal Detection Theory was initially formulated by Peterson, Birdsall, and Fox (1954) and extended by Tanner and Swets (1954) and Green and Swets (1966). Abdi (2009) extended SDT beyond the literal interpretation of physical parameters into the domains of abstract or metaphorical signals, which is pertinent to the current context. An early applications of Signal Detection Theory was to model human operator performance during target detection tasks on early radar displays. These devices suffered high levels of noise in relation to the relatively weak signal strength of the target, making the detection task difficult and probabilistic. In such situations, the radar operator and the radar each can have two states, resulting in four possible SDT system combinations:

1. A target is present on the display (a *Signal*), and it is detected by the operator – a *Hit*
2. A target is present, and is not detected – a *Miss*
3. No target is present, but one is detected (i.e. noise is mistaken for the target) – a *False Alarm* (FA)
4. No target is present, and none is detected – a *Correct Rejection* (CR)

In the following discussion, the meanings of Hit, Miss, FA, and CR are to be interpreted in the SDT context. The Hit and CR states represent the ideal operation of the system, and they may have associate *benefits*. Conversely, misses and FAs are undesirable, and each has an associated *cost*. For example, the result of a Miss could be the destruction of one's vessel by a hostile party. Equally, an FA could result in the destruction of an innocent (non-target) party by our weapon system.

The final variable is the *Decision Criterion* adopted by the operator, which defines the operator's *Response Bias*. The response bias of a *risky* operator results in more detected signals, leading to greater numbers of hits and accompanying FAs. Conversely, a *conservative* operator would incur more Misses but fewer FAs. A hypothetical unbiased *ideal* operator sets the decision criterion in an optimum manner to

minimize undesirable Misses and FAs. Any deviation from the ideal threshold represents a conservative or liberal bias.

The Response Bias adopted by the operator is also affected by the likelihood of a signal being observed, as well as by the costs and benefits of acting on a perceived signal. For example, if the likelihood of a signal and the cost of a Miss are both high (e.g. an incoming missile is both likely and catastrophic), then the operator would be very likely to perceive every stimulus as a target. On the other hand, if the cost of an FA were high (e.g. downing a civilian airliner), and the target probability extremely low (e.g. in peacetime), the operator would be unlikely to respond unless the target signal was overpowering.

The importance of SDT is that it allows an exact calculation of the optimum Response Bias, given known likelihoods of observing a signal, and with defined costs and benefits (Wickens, 1992, p. 29). This is a striking conclusion that forms the link between SDT and the aeronautical certification domain, with its highly probabilistic foundations.

SDT Definitions

At any given moment, time t, the signal can have one of two states:

Condition C_0 – the signal is absent;

Condition C_1 – a signal is present.

The system produces output data, $x(t)$, corresponding to the signal state. The operator will act on this data to make one of the following decisions or judgments:

Decision D_0 – the signal is absent;

Decision D_1 – a signal is present.

Accordingly, SDT yields four possible system states, defined as follows:

D_1C_1 – *Hit;*

D_0C_1 – *Miss;*

D_1C_0 – *False Alarm;* and

D_0C_0 – *Correct Rejection.*

Let:

$P(C_0)$ be the a-priori probability of event C_0 and

$P(C_1)$ be the a-priori probability of event C_1, then:

Events C_1 and C_0 are complementary, so $P(C_0) = 1 - P(C_1)$.

In practice, the absolute probabilities $P(D_1C_1)$, $P(D_0C_1)$, $P(D_1C_0)$, and $P(D_0C_0)$ are usually unknown, so conditional probabilities are substituted for the four system states identified above:

$P_H = P(D_1|C_1)$ is the *Hit probability,*

which is the conditional probability of D1, given that C1 has occurred. Similarly:

$P_M = P(D_0|C_1)$ is the *Miss probability.*

$P_M = 1 - P_H$, because these are the only two possible outcomes, given the *presence* of a signal.

Also:

$P_{FA} = P(D_1|C_0)$ is the *FA probability*;

$P_{CR} = P(D_0|C_0)$ is the *CR probability*.

Again, $P_{CR} = 1 - P_{FA}$, because these are the only two possible outcomes, given the *absence* of a signal.

SDT Costs

In SDT, there are two possible failure outcomes: Miss and False Alarm. These generally have different negative consequences, depending on the real-world situation. For this reason, SDT introduces two corresponding relational error *costs*:

c_{01} – *Miss Cost;*

c_{10} – *FA Cost.*

SDT Average Risk

Combining these concepts, SDT characterizes the *average risk* value of the system as:

$$\mathbf{R = c_{01}P_M P(C_1) + c_{10}P_{FA}P(C_0)} \quad (1)$$

If all the values in (1) are known, the *Bayes Criterion of Minimum Average Risk* R (R → min) yields an Optimal Detection Criterion that will maximize the system's Hits and minimize the False Alarms (Van Trees, 2001). Note that the optimum performance of the system does not eliminate Misses and FAs, because of the probabilistic nature of the system, but the Bayes Criterion does provide the optimum theoretical system performance. The only drawback of the Bayes approach is that the variables in (1) are not usually known. Nevertheless, the equation can be used as a starting point for the application of SDT for certification purposes. Before making this transition, it is necessary to examine the probabilistic underpinnings of current certification approaches.

Failure Conditions, Failures and Errors

Aeronautical Circular 23.1309-E (FAA, 2011) defines the following terms:

Error

An omission or incorrect action by a crewmember or maintenance personnel, or a mistake in requirements, design, or implementation.

Failure

An occurrence that affects the operation of a component, part, or element such that it can no longer function as intended (this includes both

loss of function and malfunction). Note: Errors may cause failures but are not considered failures.

Failure Condition

A condition having an effect on either the airplane or its occupants, or both, either direct or consequential, which is caused or contributed to by one or more failures or errors considering flight phase and relevant adverse operational or environmental conditions or external events. Figure 2 of 23.1309-E (FAA, 2011, p. 23) places maximum bounds for different Failure Condition severity levels as follows:

$$\mathbf{P(FC_i) < M_i,\ i = 1, \ldots, 4} \quad (2)$$

where:

FC_1 is a *Minor Failure Conditions*,

FC_2 is a *Major Failure Conditions*,

FC_3 is a *Hazardous Failure Conditions*,

FC_4 is a *Catastrophic Failure Conditions*, and

M_i are the Maximum acceptable values corresponding to each Failure Condition severity level.

AC 23.1309-E states:

> The probability of a failure condition occurring on an "average flight" should be determined by structured methods (see ARP 4761 for various methods) and should consider all elements (e.g., combinations of failures and events) that contribute to a failure condition. If there is only an effect when failures occur in a certain order, the calculation should account for the conditional probability that the failures occur in the sequence necessary to produce a failure condition (FAA, 2011, p. A3–1).

In particular, if a Failure Condition may be caused by 1 of n mutually exclusive failures $F_1, \ldots, F_n$, then:

$$\mathbf{P(FC) = P(FC|F_1)P(F_1) + \ldots + P(FC|F_n)P(F_n)} \quad (3)$$

This is the certification equivalent of SDT equation (1) above. In the context of an information system, (2) does not differentiate between the different costs associated with Loss-of-Function failures (*Misses*) from Hazardously Misleading ones (*False Alarms*). Equation (3) also takes no account of the potential benefits of the optional system, as there are no *benefits* terms in the equation. Accordingly, an optional safety system, of the type being addressed by this paper, might be deemed uncertifiable, despite overwhelming potential benefits. This shortcoming can be addressed by mapping the SDT approach to the certification domain.

Mapping SDT and Aircraft Certification Terms

The SDT concepts of Signals, Hits, Misses, FAs, CRs, and System Average Risk can be applied to optional aircraft safety systems, whereby a Signal is viewed as a pilot error, and a Hit is viewed as a Save by the safety system in question. Using this approach, the SDT definitions can be mapped to the certification environment as follows:

Signal represents an unaided pilot error when the safety system is not installed that can cause an accident (i.e. UPE - an unaided pilot error). An optional safety system is therefore analogous to a Signal Detection System in SDT. The associated *Signal* probability is denoted by P_{UPE}.

Hit denotes a "save" by a correctly functioning safety system, which prevents the pilot from making an error that would otherwise have been committed. An SDT *Hit* maps to a certification Save, with a probability of P_{Save}.

Miss denotes a safety system's failure to prevent an error under UPE conditions. Let's denote *Miss* by NSave (No Save) and the Miss probability by P_{NSave}.

Correct Rejection reflects the correct operation of the system in the absence of any pilot error.

False Alarm represents a safety system failure that results in Hazardously Misleading (HM) data being presented, in the absence of a UPE. The equivalent *False Alarm* probability is P_{HM}.

Miss Cost denotes the conditional probability of a Failure Condition of a specified severity level arising as a result of a safety system's failure to Save P(FC|NSave). This parameter broadly characterizes the severity of the consequences of the safety system's failure.

False Alarm Cost denotes the conditional probability of a Failure Condition of a specified severity level arising as a result of a safety system's issuing a False Alarm (or Hazardously Misleading Information) P(FC|HM). This parameter captures the severity of the consequences of the safety system's issuing a false alarm.

Applying these mappings of SDT terms, the Average System Risk from (1) can be rewritten as:

$$\mathbf{R = P(FC|NSave)P_{NSave}P_{UPE} + P(FC|HM)P_{HM}(1 - P_{UPE})} \quad (4)$$

Any possible failure in SDT can be categorized either as a Miss (NSave) or a False Alarm (HM), which are mutually exclusive, so, according to (3), the R in (4) is analogous to $P(FC_i)$ in the Certification Requirement (2) above.

The System Efficiency Concept

In (4) above, R is the risk of a Failure Condition when the system is present. We now define R_w as the risk of the same Failure Condition without the optional system. It follows that the system is *effective* if the overall risk with the system is lower than the risk without the system installed:

$$\mathbf{R < R_w} \quad (5)$$

This is the key formula for determining any safety system efficiency.
The percentage efficiency of a safety system can be considered as:

$$\mathbf{Eff(\%)= 100(R_w - R)/R_w} \quad (6)$$

Let $P_w(FC|UPE)$ be the conditional probability of specified Failure Condition without the system, under a given UPE condition. The overall risk of the specified Failure Conditions is therefore:

$$\mathbf{R_w = P_w(FC|UPE)P_{UPE}} \quad (7)$$

Using (4) and (7), we can rewrite the efficiency requirement (5) as:

$$\mathbf{P(FC|NSave)P_{NSave}P_{UPE} + P(FC|HM)P_{HM}(1 - P_{UPE}) < P_w(FC|UPE)P_{UPE}} \quad (8)$$

NSave represents a failure event under a given UPE condition. This is no different than the situation where a pilot has made an error without the system installed, so $P(FC|NSave) = P_w(FC|UPE)$, and (8) can be rewritten as

$$P_w(FC|UPE)P_{NSave}P_{UPE} + P(FC|HM)P_{HM}(1–P_{UPE}) < P_w(FC|UPE)P_{UPE}$$

Rearranging:

$$P(FC|HM)P_{HM}(1 – P_{UPE}) < P_w(FC|UPE)P_{UPE}(1 – P_{NSave})$$

By definition: $(1 – P_{NSave}) = P_{Save}$ giving our final requirement:

$$\mathbf{P(FC|HM)P_{HM}(1–P_{UPE}) < P_w(FC|UPE)P_{UPE}P_{Save}} \quad (9)$$

This formula quantitatively defines the threshold criterion at which the optional safety system statistically breaks even with the baseline unmodified aircraft, taking into account both the risks and the potential benefits of the system. The application of the preceding criterion is best illustrated using a case study.

Case Study

The following example pertains to the presentation of an aircraft position spotter during flight on an EFB-hosted electronic chart, which is currently prohibited unless an "…installed primary flight display, weather display, or map display also depict(s) own-ship position" (FAA, 2017, p. 15).

The use of a spotter undoubtedly confers some operational and safety benefits, but at the risk of misleading the crew if a software failure leads to a hazardously misleading (HM) condition. This could arise if the spotter is shown in the wrong position or orientation. The situation would result in a failure condition if the pilot(s) follow the bad data, Air Traffic Control doesn't catch the error, etc. These probabilities can be estimated and applied to (9) to quantitatively determine if the spotter confers a positive safety benefit. To do so, we introduce the following events for illustrative purposes only:

ErrSw - an untrapped software error has caused a hazardously misleading (HM) condition, in the absence of a prevailing pilot error (i.e. no UPE);

ErrGPS - incorrect GPS or navigation input to the spotter has caused a hazardously misleading (HM) condition, in the absence of a prevailing pilot error (i.e. no UPE);

ErrDB - a chart database error has caused a hazardously misleading (HM) condition, in the absence of a prevailing pilot error (i.e. no UPE);

Nr - "Not recognized": the crew fail to recognize the HM that the aircraft is not at the displayed position;

E_1 - the HM information is in a dangerous sense (e.g. the error biases the crew towards an occupied runway, rather than away from it);

E_2 - the crew follows the HM information, despite other visual or navigation cues;

E_3 - Air Traffic Control fails to detect the hazardous maneuver; and

E_4 - the crew action actually causes an accident (e.g. collision with terrain, obstacles, or another aircraft).

Using the definitions above, the combined probability of Hazardously Misleading Information from the three identified causes is:

$$\mathbf{P_{HM} = 1 - (1-P_{ErrSw})(1-P_{ErrGPS})(1-P_{ErrDB})} \quad (10)$$

Nr, E_1, E_2, E_3, and E_4 are the necessary events following HM that will lead to a Failure Condition, so:

$$\mathbf{P(FC|HM) = P(E_4E_3E_2E_1|NrHM)P(Nr|HM)} \quad (11)$$

Once the crew has failed to recognize a hazardously misleading spotter event, the probability of the subsequent events (E_1-E_4) leading to a Failure Condition are identical, whether the system is present or not. For example: ATC is no more or less likely to detect a deviation caused by an HM-induced spotter-error than one caused by an unaided pilot error without the system installed. Similarly, the likelihood of a random unaided pilot error (UPE) being in a dangerous sense is identical to the probability

that a random HM software error is also in a dangerous sense. For example: random software and pilot errors would be expected to have equal probabilities of biasing the crew towards, or away from, an occupied runway. Summarizing this concept:

$$\mathbf{P(E_4E_3E_2E_1|NrHM) = P_w(FC|UPE)} \quad (12)$$

Using (11) and (12), the system efficiency criterion (9) can be rewritten as:

$$P_w(FC|UPE)P(Nr|HM)P_{HM}(1–P_{UPE}) < P_w(FC|UPE)P_{UPE}P_{Save}$$

Simplifying:

$$\mathbf{P(Nr|HM)P_{HM}(1–P_{UPE}) < P_{UPE}P_{Save}} \quad (13)$$

This formula makes an interesting contrast with the standard certification requirement we saw in (2): $P(FC_i) < M_i$, $i = 1, \ldots, 4$.

For the electronic chart spotter, formulas (10) and (13) can be combined to calculate the maximum acceptable probability of an untrapped software error leading to an HM event:

$$\mathbf{1–(1–P_{ErrSw})(1–P_{ErrGPS})(1–P_{ErrDB}) < P_{Save}P_{UPE}/(P(Nr|HM)(1–P_{UPE}))} \quad (14)$$

Formula (14) yields a quantified measure of the required system reliability. The final step in the analysis is to examine the variation of the maximum allowable probability of a Hazardously Misleading software error P_{ErrSw}. This is best visualized graphically, and requires the introduction of three final constructs.

P_{Save} and P(Nr|HM) in (14) are difficult to calculate with absolute accuracy, but a solution can be derived by revisiting the automobile airbag example used in the introduction. It is doubtful that accurate figures could be derived for airbag "saves" and for "losses" caused by malfunctioning airbags. Nevertheless, the order of magnitude of the save ratio can be estimated. The same analogy applies to the un-quantified probabilities above: the order of magnitude of the ratio $P_{Save}/P(Nr|HM)$ can be estimated, with sufficient accuracy for this analysis. This ratio is used as abscissa for the required system reliability plot.

Similarly, the effect of a wide range of P_{UPE} values should be examined in order to determine the system sensitivity to the probability of Unaided Pilot Errors. For this reason, (14) is used to produce a family of curves for varying P_{UPE} values. These have been bounded within a range of range 10^{-2} -10^{-4} because the former would represent many thousands of errors every day, when viewed across all flight operations worldwide. Conversely, the latter would imply that a representative 20,000-hour pilot has only made one such error in his or her career, based on an average stage-length of two hours.

The last assumption relates to the values for P_{ErrGPS} and P_{ErrDB}**.** These are effectively the probabilities of a hazardously misleading GPS

position (independent of the EFB and its software), and of a hazardous chart database error. These probabilities can be derived in a number of ways, including service history and FOQA data reviews, but for the purpose of the case study, they are arbitrarily assigned the following values:

$P_{ErrGPS} = 10^{-6}$

$P_{ErrDB} = 10^{-5}$

Figure 1 illustrates the result of applying the preceding assumptions to (14). The y-axis (max. acceptable P_{ErrSw}) has a logarithmic scale with the inverse order of values.

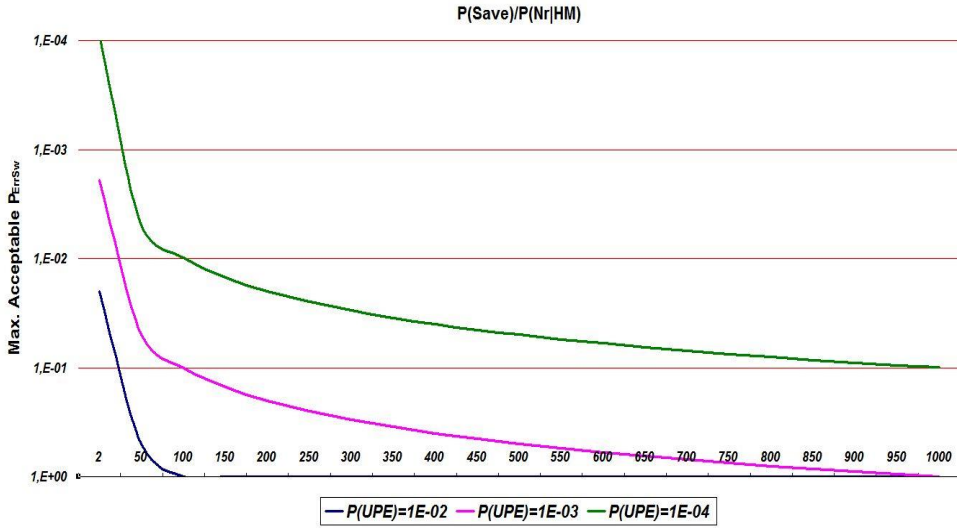

Figure 1. Minimum Acceptable Software Reliability as a Function of System Benefit/ Risk Analysis.

The following example illustrate the practical application of Figure 1. Assuming that the pilot's probability of making an unaided error is 10^{-3}, and that the system saves 25 times as often as it hazardously misleads, then the required system reliability to achieve a net beneficial effect is approximately 2.5×10^{-2}. This corresponds to DO-178C Level D software (RTCA, 2012), which is achievable by Commercial-Off-The Shelf (COTS) products and applications. If this performance requirement is exceeded, the optional system would yield a positive safety improvement over the baseline, even though the assumed reliability is several orders of magnitude below that required for navigation systems.

Figure 1 also shows that the software reliability must be increased (i.e. the failure rate must decrease), when either of the following occurs:

1. The pilot becomes *more* reliable; and/or

2. The system Save/Unrecognized Hazard ratio decreases.

Conclusions

Signal Detection Theory and Bayesian optimization methods can be applied to the certification of optional aircraft systems, and a formal method has been developed that allows the numerical optimization of the risk/benefit ratio of such systems. Using representative data from the case study of a spotter on an electronic chart, it has been demonstrated that safety benefits would be achieved, even with the software reliability levels typically associated with COTS software such as Windows™ which are significantly below the current certification standards. The method makes few domain assumptions, and is based on the underpinnings of SDT and Bayesian probability theory, with well-established validity and reliability. Accordingly, the technique should have broad application to the certification of all optional aircraft systems.

Nomenclature

ATC – Air Traffic Control
CR – Correct Rejection
E_1 - the HM information is in a dangerous sense (e.g. the error biases the crew towards an occupied runway, rather than away from it)
E_2 - the crew follows the HM information, despite other visual or navigation cues
E_3 - Air Traffic Control fails to detect the hazardous maneuver
E_4 - the crew action actually causes an accident (e.g. collision with terrain, obstacles, or another aircraft)
ErrDB - a chart database error has caused a hazardously misleading (HM) condition, in the absence of a prevailing pilot error (i.e. no UPE)
ErrGPS - incorrect GPS or navigation input to the spotter has caused a hazardously misleading (HM) condition, in the absence of a prevailing pilot error (i.e. no UPE)
ErrSw - an untrapped software error has caused a hazardously misleading (HM) condition, in the absence of a prevailing pilot error (i.e. no UPE)
FA – False Alarm
FAA – Federal Aviation Administration
FC – Failure Condition
FOQA – Flight Operations Quality Assurance
HM – Hazardously Misleading
Nr - "Not recognized": the crew fail to recognize the HM that the aircraft is not at the displayed position
NSave – No Save
P – Probability
R – Risk of a failure condition when the system is present
R_w – Risk of the same failure condition Without the system
RTCA - Radio Technical Commission for Aeronautics
SAE – Society of Automotive Engineers
UPE - Unaided Pilot Error

References

Abdi, H. (2009). Signal detection theory. In B. McGaw, P. L. Peterson, & E. Baker (Eds.), *Encyclopedia of Education* (3rd ed., pp. 1-10). New York, NY: Elsevier.

Federal Aviation Administration. (2011). *Advisory Circular 23.1309-E, System safety analysis and assessment for Part 23 airplanes*. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_23_1309-1E.pdf

Federal Aviation Administration. (2017). *Advisory Circular 120-76D, Authorization for use of electronic flight bags.* Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-76D.pdf

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Peterson, W. W., Birdsall, T. G., & Fox, W. (1954). The theory of signal detectability. *Information Theory, Transactions of the IRE Professional Group On, 4(4),* 171-212. doi:10.1109/TIT.1954.1057460

RTCA. (2000). *Design assurance guidance for airborne electronic hardware* (DO-254). Washington, DC: RTCA, Inc.

RTCA. (2012). *Software considerations in airborne systems and equipment certification* (DO-178C). Washington, DC: RTCA, Inc.

SAE. (2010). *ARP4754, Guidelines for development of civil aircraft and systems*. Retrieved from www.sae.org/technical/standards/arp4754a.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*(6), 401-409. doi:10.1037/h0058700

Van Trees, H. L. (2001). *Detection, estimation and modulation theory.* New York, NY: John Wiley & Sons, Inc.

Wickens, C. D. (1992). *Engineering psychology in human performance* (2nd ed.). New York, NY: HarperCollins Publishers Inc.