Know the SCOR for Multifactor Strategy of Experimentation: Screening, Characterization, Optimization and Ruggedness Testing

Mark J. Anderson, PE, CQE, MBA
Stat-Ease, Inc. Minneapolis, MN

By way of example, this article lays out a strategy for design of experiments (DOE) that provides maximum efficiency and effectiveness for development of a robust system. It broadens the scope of a prior article (Anderson and Whitcomb 2014) that spelled out how to right-size multifactor tests via statistical power-calculations—a prerequisite for DOE success.

Introduction

In the 1950s, when George E. P. Box developed response surface methods (RSM) as the keystone tool for design of experiments (DOE) (Box and Wilson 1951), a sequential approach to experimentation emerged as a winning strategy for process development (Snee 2009). The main elements of this strategy—laid out in Figure 1—are screening, characterization, optimization, and ruggedness testing or "SCOR". SCOR scores big by taking small steps—putting no more than 25 percent of the total resources into any single experiment, thus allowing for changes in direction all along the way. It is fast, flexible, and statistically rigorous.

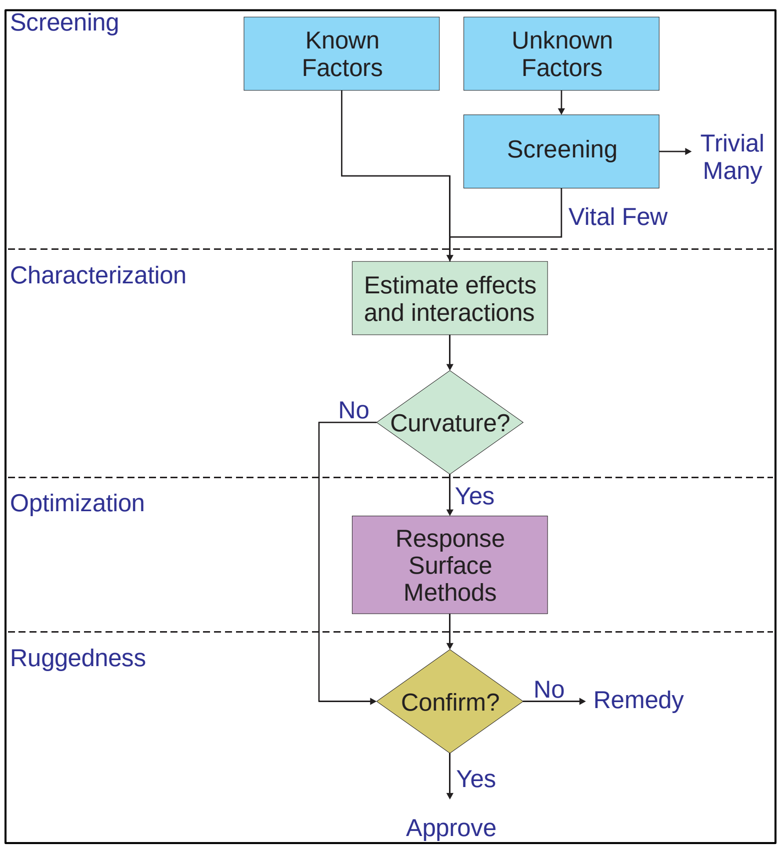

Figure 1: A Flowchart for a Winning Strategy of Experimentation

This now-proven strategy of experimentation begins with broad, but shallow, fractional experiment designs that screen previously untested factors. Be careful at this stage not to cut the runs back so far that main effects become aliased with two-factor interactions. Another word of caution—do some range-finding on factors and their levels: Being bold will pay dividends, but not if unsafe. As shown in Table 1, Resolution IV standard two-level fractional-factorials (designated "2^{k-p}", where k is the number of factors and p the fraction) provide the stoutest alternative for screening. However, if runs must be reduced and they can be spared without undermining power, minimum-run resolution IV (MR4) designs, also two level, come to the forefront. Another alternative (not shown on the table) is the definitive screening design (DSD), which, for better (providing information on curvature) or worse (the operational bother of more settings), requires three levels of every factor (Anderson and Whitcomb 2015).

In any case, during this screening phase experimenters seek to discover the vital few factors that create statistically significant effects of practical importance. As a general rule, only 20 percent of the factors screened will merit further investigation, leaving the other 80 percent—the trivial many—behind.

After throwing out the unimportant factors, preferably by holding them fixed or blocking them out, the SCOR program enters the phase of characterizing interactions. This requires higher-resolution (V or better, or MR5), or full, two-level factorial designs. Keep in mind that whereas traditional one-factor-at-a-time (OFAT) approaches can detect main effects, albeit inefficiently, OFAT cannot detect interactions. Two-factor interactions (2FIs) cannot be ignored—mastering these 2FIs often proves to be the key to success.

At this characterization point for SCOR, performance may be nearing peak levels, thus it pays to put in three or four center points (all factors set at mid-levels) to check for curvature. If mission-critical curvature is detected, the time comes to deploy response surface methods (RSM) via the central composite design (CCD), Box-Behnken design (BBD), or computer-generated optimal designs (Anderson and Whitcomb 2016).

If the experimenters are skilled and lucky, they will succeed immediately at the screening stage, or more often, after accomplishing characterization. Perhaps an optimization experiment will be required to accomplish the objective. However, no matter how the experimental program reaches its goal—even without DOE whatsoever, running a multifactor ruggedness test provides the ultimate confirmation. Ruggedness testing is a special application of a statistically designed experiment that examines a great many field conditions to determine which, if any, might affect the system (Anderson 2017). Plackett-Burman designs provide maximum utility for ruggedness testing due to their simplicity (only two levels), flexibility (templates in multiples of four), and efficiency (saturated arrays allowing up to 1 fewer factor than the number of runs, for example, 11 variables in only 12 runs) (Haase 2011).

As noted in Table 1, low-resolution (III) standard fractional factorials (2^{k-p}) also serve well for ruggedness tests. If all goes well, the system flies through all combinations of potentially upsetting factors with nary a hitch, that is, no effects of any practical importance for their setback on performance. However, if the test fails, a second set of runs with all levels opposite of the first—called a fold over design—can resolve the main effects and interactions of concern (Anderson et al., 2018).

Table 1: Experiment-design Choices

Phase	2^{k-p}_{III}	2^{k-p}_{IV}	2^{k-p}_{V}	2^{k}_{Full}	MR4	MR5	PB	CCD	BB
Screen	No	✓	OK	OK	✓	OK	No	No	No
Characterize	No	No	✓	✓	No	✓	No	OK	OK
Optimize	No	No	No	No	No	No	No	✓	✓
Ruggedness	✓	OK	No	No	OK	No	✓	No	No

Optimal designs are not shown in Table 1 due to them being beyond the scope of this overview. They can be customized by DOE software for any phase of SCOR. Also, this article does not discuss DOE tools for mixtures geared to formulators of materials (metals, plastics, composites, adhesives, coatings, etc.) with components that must be constrained to a fixed total (e.g., 100 weight percent). However, a similar SCOR strategy can be applied (Anderson, Whitcomb and Bezener 2018).

SCOR Strategy by Example

To illustrate the SCOR strategy of experimentation, let's work through the hypothetical development of a welding process and the testing of the final part, an example that falls within the province of military application, as in the manufacture of armor (DOD 1998).

Figure 2: An Aviation Structural Mechanic Making a Practice Weld (Naval Air Systems Command 2018)

Being the weak point mechanically, the welds must exhibit high tensile strength—above anything ever achieved before: more than 50,000 psi. This being new territory for the engineering team, they brainstorm many new process factors. After much discussion, the team narrows down the field to 11 factors, of which 2 are known to create substantial effects—current and metal substrate. The effects, if any, created by the other 9 factors remain unknown. While setting aside the known factors, these potentially new variables, listed below, are evaluated via a screening design:

A. Angle, degrees: 60 - 80
B. Substrate thickness, millimeters (mm): 8 - 12
C. Opening, mm: 1.5 - 3
D. Rod diameter, mm: 4 - 8
E. Rate of travel, mm/second: 0.5 - 2
F. Drying of rods, hours: 2 - 24
G. Electrode extension, mm: 6 - 15
H. Preheating Temperature, degrees F: 250 - 350
J. Edge prepped: No - Yes

The labeling of factor 9 (Edge Prep) as “J” is not a typo—the letter “I” being reserved for the intercept for DOE modeling.

The experimenters consider three designs:

1. Study only the first 7 factors in an 8-run standard fractional factorial (2^{7-4})—resolution III.
2. Screen all 9 factors in the 32-run standard resolution IV fraction (2^{9-4}).
3. Choose an MR4 design with 2 extra runs to allow for a few being botched.

Option 1 is quickly rejected for ignoring two of the possibly important factors and, also, for it being badly aliased (main effects confounded with 2FIs). The second option requires more runs than the experimenters can expend. Option 3, which requires only 20 runs, yet avoiding the aliasing of main effects with two-factor interactions, hits the spot. It’s laid out in Table 2 by standard order (actual runs were randomized) with the results for tensile strength of the welds reported.

Table 2: Screening Design

#	A	B	C	D	E	F	G	H	J	Tensile Strength psi
1	80	8	3	8	0.5	24	6	350	No	43880
2	80	8	1.5	8	0.5	24	15	250	No	46100
3	60	12	1.5	8	0.5	24	6	250	Yes	46770
4	80	12	3	8	0.5	2	15	250	Yes	51290
5	60	8	3	8	0.5	2	6	250	No	43340
6	60	8	1.5	4	0.5	2	15	350	Yes	44250
7	80	12	1.5	4	0.5	2	6	250	No	48890
8	80	8	3	4	0.5	24	6	250	Yes	45810
9	80	8	1.5	8	2.0	2	6	250	Yes	47060
10	80	12	1.5	8	0.5	2	6	350	Yes	51650
11	60	12	1.5	4	2.0	2	15	250	Yes	46260
12	60	12	3	4	2.0	2	6	350	Yes	46810
13	80	8	3	4	2.0	2	15	350	No	43670
14	60	8	1.5	4	2.0	24	6	350	No	44470
15	80	12	1.5	4	2.0	24	15	350	Yes	51200
16	80	12	3	8	2.0	24	6	250	No	49080
17	60	8	3	8	2.0	24	15	350	Yes	46770
18	60	12	1.5	8	2.0	2	15	350	No	45280
19	60	12	3	4	0.5	24	15	350	No	45890
20	60	8	3	4	2.0	24	15	250	No	45040

Out of all the factors screened, only three emerge large (falling far right of the line) on the half-normal plot of effects (Figure 3): A, B (including their interaction-AB) and J.

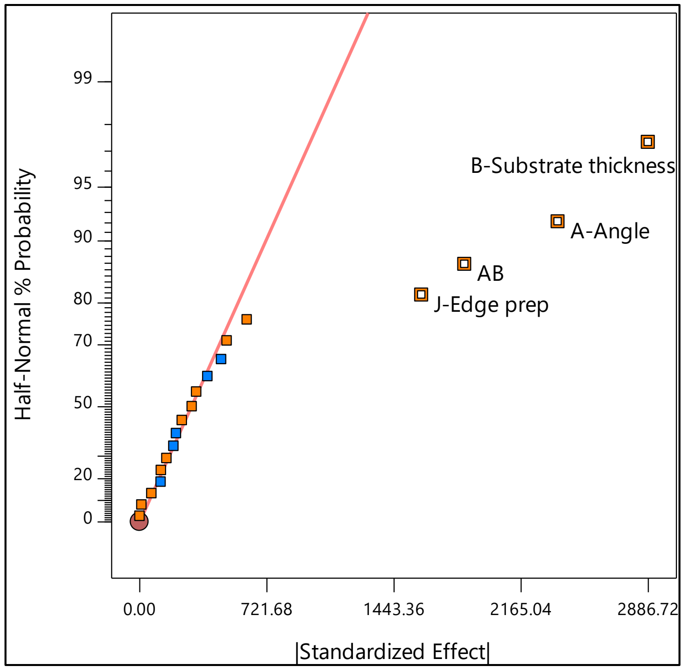

Figure 3: Factors A, B and J Stand Out for Welding

Throwing off the 8 trivial-many factors (lined up near zero effect), the experimenters carry on with the 3 vital few (A, B and J), to the next phase—characterization of interactions. As shown in Figure 4, with the 2 known factors set aside earlier now merged in, that leaves 5 factors to be studied in greater depth.

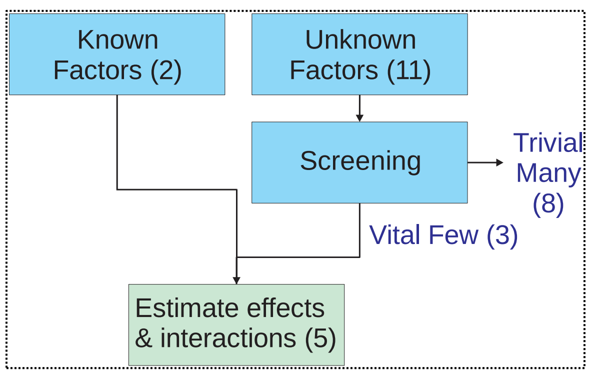

Figure 4: Screening Completed for Welding Process

Factor J—the edge prep—exhibited a main effect only, and the engineers were sure that it should be done, which proved to be the case from the screening experiment, so they decide to hold this fixed at the “yes” level. That leaves 4 factors to be studied:

A. Angle, degrees: 60 - 80
B. Substrate Thickness, mm: 8 - 12
C. Current, amps: 125 - 160
D. Metal Substrate, stainless steel: SS35 - SS41

The only design choice with high enough resolution to clearly detect two-factor interactions is the full, 16-run, two-level factorial. To test for curvature and provide measures of pure error, the experimenters add 3 center points of the numerical factors A, B, and C at each of the two categories stainless steel (D). The results shown in Figure 5 (the green triangles being the error estimates) reveal a complex model with two interactions involving three factors that affect the tensile strength of the weld.

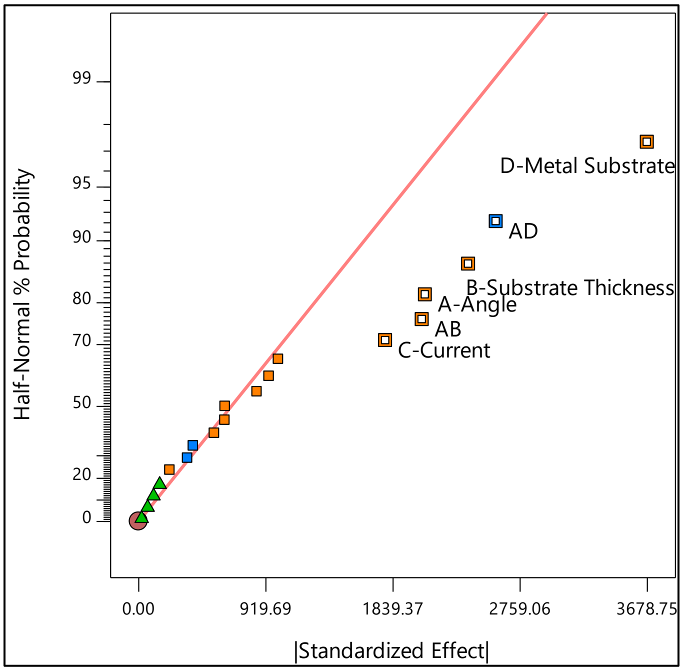

Figure 5: Characterization Study Reveals Interactions between Welding Factors

Furthermore, as the 3D plot in Figure 6 clearly shows (for the winning metal—the SS41), the curvature in response creates an appreciable lack of fit at the center of the experimental region.

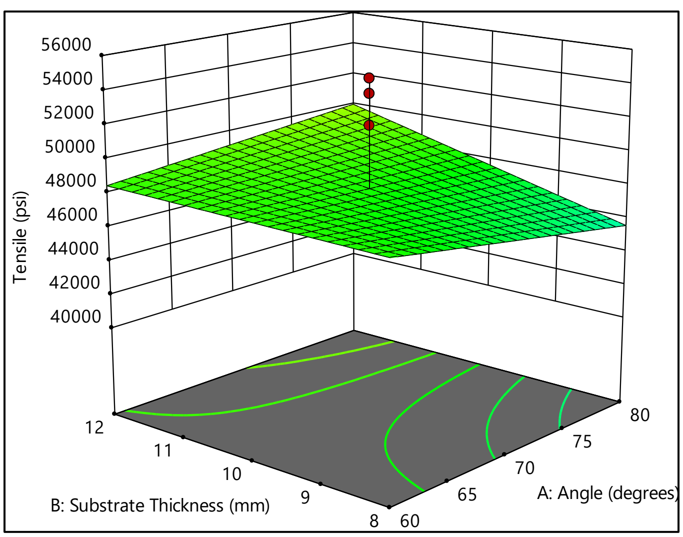

Figure 6: Pronounced Curvature!

This is a signal (clue) that optimization via RSM will be required to adequately model the tensile strength. In this case with the peak being clearly near the middle, the experimenters simply add further runs that augment the characterization factorial into a central composite design (CCD)—the standard

RSM template. To simplify matters and reduce the number of runs required to close the gap on the curvature, they eliminate the poorer-performing metal substrate (SS35) from further consideration. Table 3 shows the second block of runs with each of the three remaining factors pushed out to axial levels for leverage on estimating the nonlinearity at the center of the space.

Table 3: Central Composite Design

#	Blk	Point type	A	B	C	Tensile Strength psi
1	1	Factorial	60.0	8.0	125.0	47910
2	1	Factorial	80.0	8.0	125.0	44380
3	1	Factorial	60.0	12.0	125.0	48600
4	1	Factorial	80.0	12.0	125.0	47370
5	1	Factorial	60.0	8.0	160.0	47430
6	1	Factorial	80.0	8.0	160.0	46540
7	1	Factorial	60.0	12.0	160.0	49370
8	1	Factorial	80.0	12.0	160.0	52970
9	1	Center	70.0	10.0	142.5	51770
10	1	Center	70.0	10.0	142.5	53620
11	1	Center	70.0	10.0	142.5	54510
12	2	Axial	53.2	10.0	142.5	48850
13	2	Axial	86.8	10.0	142.5	48890
14	2	Axial	70.0	6.6	142.5	46600
15	2	Axial	70.0	13.4	142.5	50810
16	2	Axial	70.0	10.0	113.1	50460
17	2	Axial	70.0	10.0	171.9	53200
18	2	Center	70.0	10.0	142.5	53300
19	2	Center	70.0	10.0	142.5	53300

Figure 7 shows the true curvature for the surface, depicted with current (C) set to its factorial high level of 160 amps, which provides optimal results for welding—well above the goal of 50,000 psi tensile strength—when done at an angle (A) of 73 degrees over a substrate with an 11-millimeter thickness.

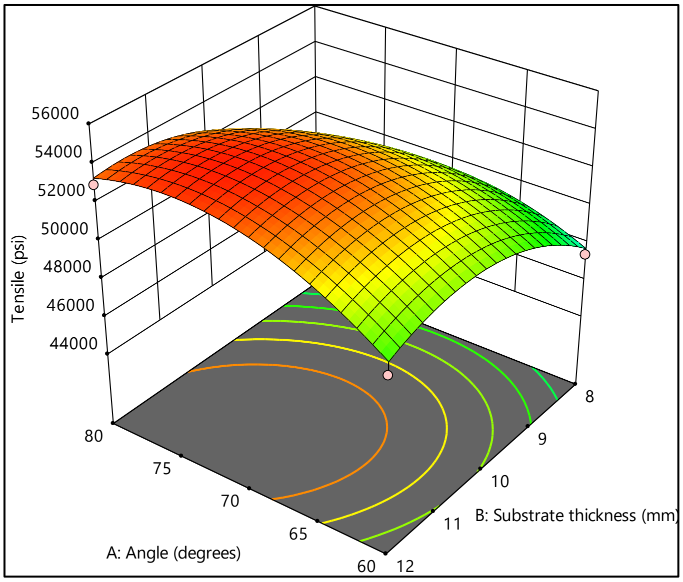

Figure 7: Ideal Setup for Welding (current at 160 amps)

A follow-up series of 6 welds at these conditions produces welds ranging well within the 95 percent model-prediction.

All that remains for achieving SCOR is to see if the welding process will be robust to production conditions by running a ruggedness test. The engineering team identifies 11 factors of concern—ambient conditions and the like. They set ranges from low (minus) to high (plus) that span the majority (95 percent or so) of the normal variation based on historical records. A 12-run Plackett-Burman design conveniently provides adequate power to detect changes in tensile strength of any importance. As shown in Figure 8, the half-normal plot of effects shows nothing significant.

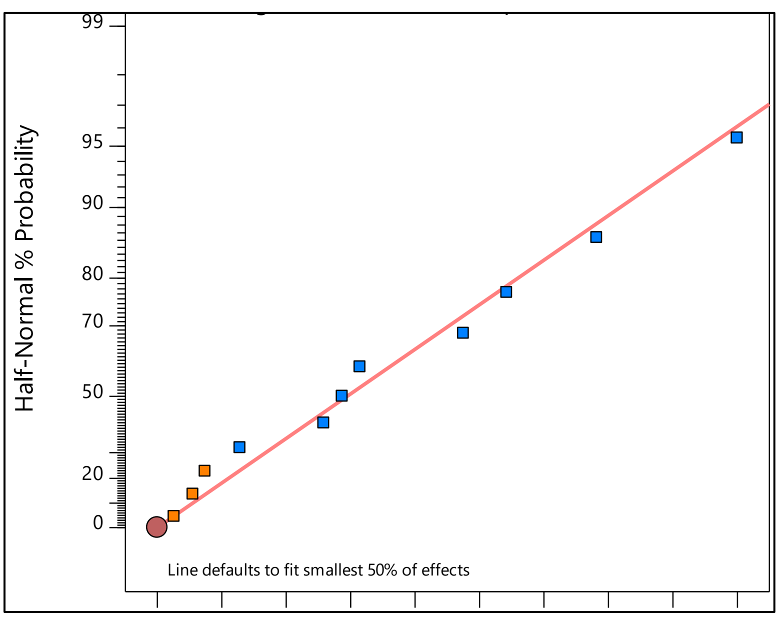

Figure 8: Half-normal Plot of Effects from Ruggedness Test

Mission accomplished.

Conclusion

The SCOR strategy provides a tried-and-true path to process improvement via an iterative series of statistically designed experiments. If powered properly by sufficient runs (sample size), it cannot fail to be productive whether it meets mission objectives or not. Because a full SCOR maps out regions that fail, it will always provide great value by the process of elimination, even when coming up short of a successful outcome.

By breaking down the research program into small steps, SCOR allows experiments to react to results along the way, thus reducing wasteful runs. For example, in the welding case, testing all 11 factors in a single RSM design would have required 96 runs for a CCD (or 88 runs for a minimal optimal design). The sequential SCOR experimentation required only 50 runs– 20 for screening, 22 to characterize, and 8 for RSM optimization. (In either case, the confirmation runs and ruggedness tests would have been done.)

If you know the SCOR, you will make the most of the tools of DOE.

MARK J. ANDERSON, PE, CQE, MBA is a principal and general manager of Stat-Ease, Inc. He is lead author of three books on the design of experiments: "DOE Simplified: Practical Tools for Effective Experimentation" (Anderson and Whitcomb 2016), "RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments" (Anderson and Whitcomb 2015) and "Formulation Simplified: Finding the Sweet Spot through Design and Analysis of Experiments with Mixtures" (Anderson, Whitcomb and Bezener 2018). Email: mark@statease.com

Acknowledgments

The welding case came about from a collaboration of Wayne Adams and Shari Kraber to illustrate the strategy of experimentation for students of Stat-Ease workshops. I also credit our founder Patrick Whitcomb for his pioneering work to make DOE easy for non-statistical experimenters—much of which is manifested in this article.

References:

Adams, W. 2015. "How to Resolve Partial Aliasing." *Stateaser*. September. Stat-Ease, Inc. Minneapolis.

Anderson, M. J. 2017. "Ruggedness Testing Made Easier with Graphical Effects Analysis. Smart Data: Design of Experiments." IFPAC Annual Meeting, March 2, 2017, North Bethesda, Maryland.

Anderson, M. J. et al. 2018. "Fractional Factorial with Foldover." Design-Expert® software version 11 tutorial. Stat-Ease, Inc. Minneapolis, www.statease.com/docs/v11/tutorials/foldover-tut.html.

Anderson, M. J. and P.J. Whitcomb. 2004. "Screening Process Factors in The Presence of Interactions." *The 58th Annual Quality Congress Proceedings*, American Society of Quality.

Anderson, M. J. and P.J. Whitcomb. 2014. "Employing Power to 'Right-Size' Design of Experiments." *ITEA Journal*, 35, no. 1: 40-44.

Anderson, M. J. and P.J. Whitcomb. 2015. Chapter 5-Fractional Factorials. Appendix: Minimum-Run Designs for Screening. *DOE Simplified, Practical Tools for Effective Experimentation, 3rd Edition*. New York, NY. Productivity Press.

Anderson, M. J. and P.J. Whitcomb. 2016. *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments, 2nd Edition*. New York, NY. Productivity Press.

Anderson, M. J., P.J. Whitcomb, and M. A. Bezener. 2018. *Formulation Simplified: Finding the Sweet Spot through Design and Analysis of Experiments with Mixtures*. New York, NY. Productivity Press.

Box, G. E. P. and K. B. Wilson. 1951. "On the Experimental Attainment of Optimum Conditions." *Journal of the Royal Statistical Society. Series B (Methodological)* 13, no. 1: 1-45.

DOD. 1998. Department of Defense Handbook, Metal-Arc Welding of Homogeneous Armor, MIL-HDBK-1941.

Haase, C. L. 2011. "Tailoring the Statistical Experimental Design Process for LVC Experiments." Air Force Institute of Technology, Thesis.

Naval Air Systems Command. "FRCSW School Qualifies Navy Welders." *NAVAIR News*, Sep 12, 2018.

Snee, R. 2009. "Raising Your Batting Average" Remember the Importance of Sequence in Experimentation. *Quality Progress* 42, no.12: 64-68.