

GO FAST > THINK BIG



Context-Responsive Cost Evaluation: Dynamic Approach to Cost Estimate Reviews

ICEAA Conference 2024

By: **Brittany Holmes and Christan Yarbrough**
Missile Defense Agency
Cost Estimating Directorate
May 16, 2024

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Approved for Public Release 24-MDA-11758 (11 Apr 24)



Purpose & Agenda

Purpose: Present the Missile Defense Agency's (MDA) historical approach to cost estimate scoring, define scoring considerations, explore scoring alternatives, and consider new scoring methods

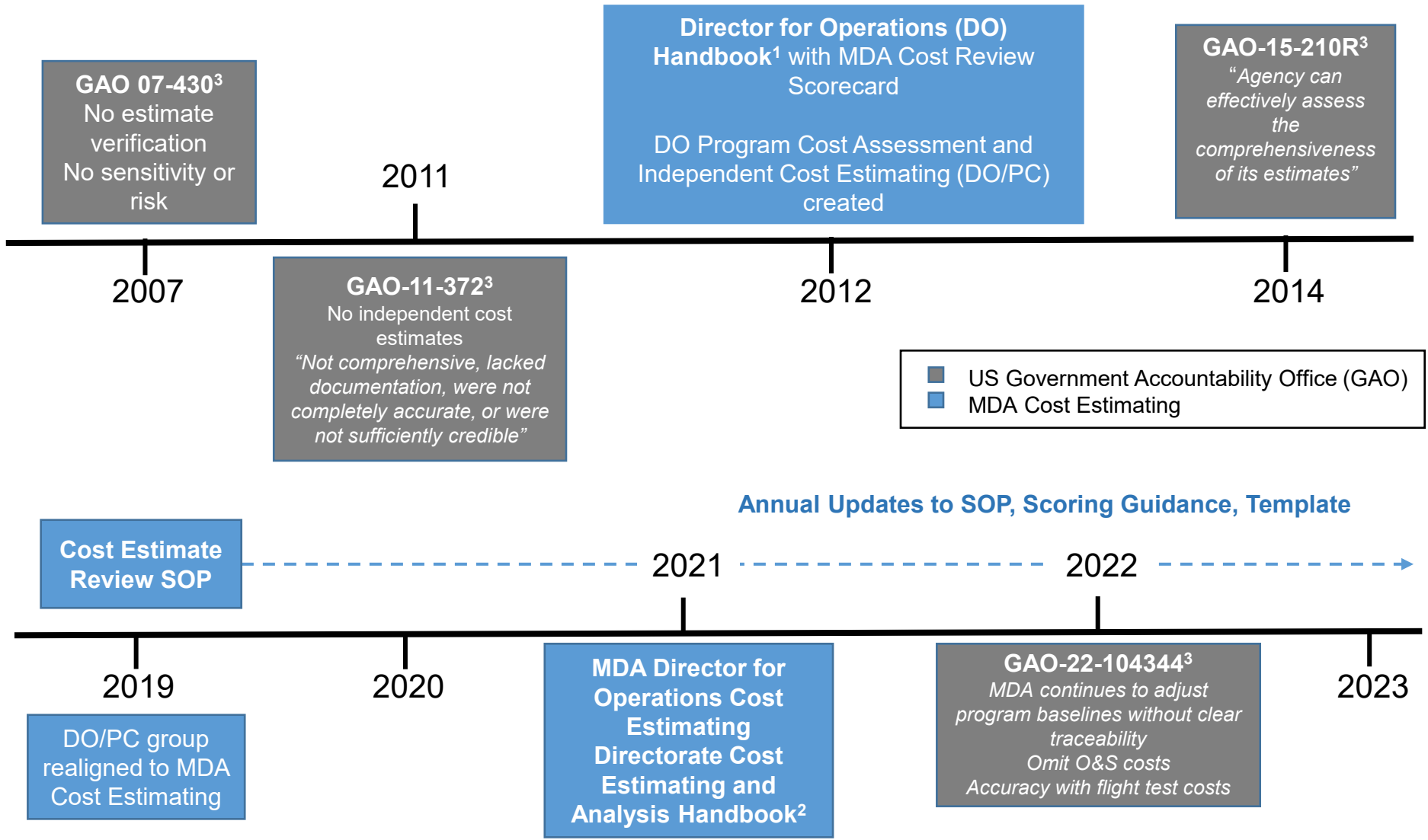
Agenda:

- **MDA's Cost Estimate Scoring**
 - History
 - Scoring Analysis
- **Scoring Considerations**
 - Goodhart's Law
 - Measures of Effectiveness (MOE)/Measures of Performance (MOP)
- **Scoring Alternatives**
 - Standards Based
 - Contract Based
 - Mastery Based
 - Competency Based
 - Specifications Based
- **Considerations for New Scoring Method**
- **Conclusions**



Cost Review Timeline

DOC – Cost Estimating Directorate



¹Approved for Public Release: 12-MDA-6903 ²Approved for Public Release: 21-MDA-10893
³GAO reports available publicly at gao.gov



MDA's Scoring Method Evolution (1 of 2)

DOC – Cost Estimating Directorate

- Handbook 2012¹ developed MDA Cost Estimate Score Card
 - Overall score is an average of the 8 criteria scores
 - Separate questions for each criteria
 - Basis in GAO Cost Estimating and Assessment Guide
- Standard Operating Procedure (SOP), 2019
 - Kept same criteria, but added more detail to each question

PROGRAM X COST REVIEW	Scoring Scale				
	Met (4)	Substantially Met (3-4)	Partially Met (2-3)	Minimally Met (1-2)	Not Met (0-1)
Comprehensive		3			
Well-Documented		3.5			
Accuracy			2		
Ground Rules & Assumptions		3.2			
Data			2.5		
Sound Methodology		3.3			
Sensitivity Analysis				0.5	
Cost Risk		3.2			
Estimate Average Score =					2.6

Figure 1.16. Example of Score Calculation for an Estimate

COST REVIEW SUMMARY	Scoring Scale				
	Met (4)	Substantially Met (3-4)	Partially Met (2-3)	Minimally Met (1-2)	Not Met (0-1)
1) Comprehensive, Accurate, & Credible					
2) Well-Documented					
3) Ground Rules & Assumptions					
4) Data					
5) Methodology					
6) Sensitivity Analysis					
7) Cost Risk					
Estimate Average Score =					

¹Approved for Public Release: 12-MDA-6903



MDA's Scoring Method Evolution (2 of 2)

DOC – Cost Estimating Directorate

- **SOP, 2020-2023**
 - **Changed from 8 criteria to 4 criteria:**
 - Well-Documented
 - Comprehensive
 - Accurate
 - Credible
 - Added and revised questions
 - Added model walk-throughs
 - Revised questions
 - Delphi methodology for review panel
 - Added pages to the mandatory template
- **Unchanged in MDA Cost Estimating Policy from 2012 to 2023 the score:**
 - “gauges the quality of an estimate”
 - “represent the maturity of an estimate”
 - “An estimate conducted early in a program’s life cycle may have lower scores due to unavailable, immature, or incomplete data”



Comprehensive Question Evolution

DOC – Cost Estimating Directorate

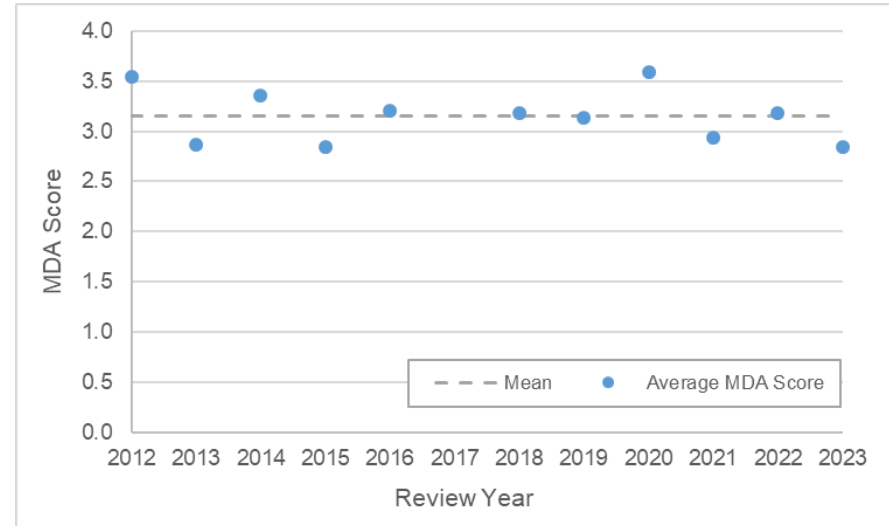
2012	2019	2020	2023	#	Question
Comp				0.0	It is comprehensive, includes all possible costs, ensures that no costs were omitted or double-counted and explains and documents key assumptions
		Comp		1.0	The estimate is comprehensive
	Comp, Acc, Cred		Comp	1.1	Supporting documentation is provided and/or available to define the estimate scope (e.g., CARD, acquisition strategy, etc.)
				1.2	It reflects the schedule at a referenced point in time
				2	Are schedules provided? Do the provided schedule(s) align to an approved requirements document? Are any changes from the requirements clearly documented?
				1.3	Ground rules and assumptions (GR&A) reflect supporting documentation or expert judgment
		Comp		2.0	The estimate captures program scope in a logical manner.
	Comp, Acc, Cred	Comp		2.1	It captures the complete technical scope outlined in the supporting documentation (e.g., CARD, acquisition strategy, etc.)
				Comp	2.2
			4		Does the work breakdown structure align to the total scope/requirements?
	Comp, Acc, Cred	Comp		3	It is a complete LCCE, accounting for development (if any), procurement, O&S (as applicable), AND disposal
		Comp		4	Identifies estimate methodologies (e.g., Analogy, Parametric, Engineering, etc)
<input type="checkbox"/> Scored <input type="checkbox"/> Presented			Comp	6	Do the cost estimates cover all scope the MDA Director is responsible for that inform MDA's annual President's Budget (PB) and Program Objective Memorandum (POM) processes (as directed per DOC review SOP)?

Comprehensive (Comp), Accurate (Acc), Credible (Cred)



Score Analysis

- **Eight MDA programs with scores between 2012 to 2023**
 - Excludes new programs, programs with one score, and functional groups (ex. Engineering and Test)
 - Tested statistical relationships of scores and
 - # Pages in template
 - Scoring format
 - Average age of the program
 - Annual reported baselined values in BY2023\$
 - DOC Director
 - Team Lead
 - **Weak negative correlations**
 - # of Pages and Score
 - Age and Score
 - **Average score fluctuates around an average (3.2) from 2012-2023**





Scoring Analysis Conclusions

DOC – Cost Estimating Directorate

- **Score not reflecting cost estimates quality attributes as intended**
 - Expect ongoing program scores to increase over time
- **Possible reasons:**
 - Estimate quantity increasing
 - Team turn-over
 - MDA Program requirement fluctuations
 - Cost review requirement fluctuations
 - Moving standard (last year's 4.0 is the new 3.5)
 - Program to program comparison vice program to standard comparison
 - Scoring panel composition
 - Larger panel, higher scores
 - Some panel members 30% easier than average rater
- **Scoring strategy should be revised considering**
 - Goodhart's Law
 - New grading mechanisms

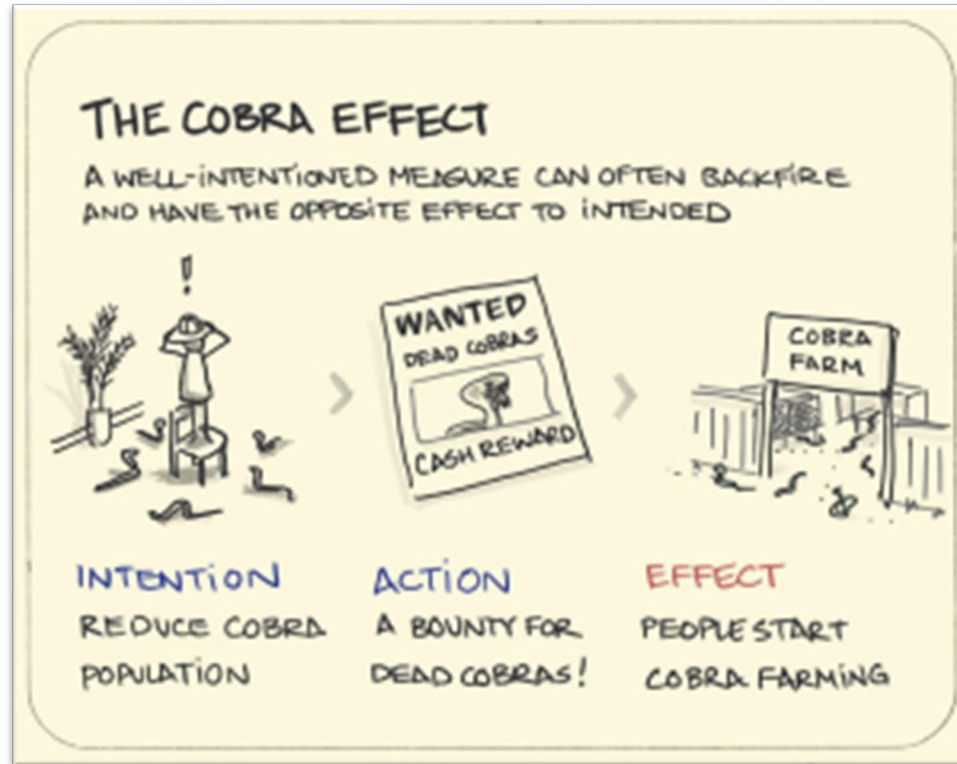


What is Goodhart's Law?

- **1975 Charles Goodhart in paper examining the relationship between money supply and inflation:**
 - “Any observed statistical regularity will tend to collapse once pressure is place upon it for control purposes.”
- **1979 Donald Campbell in paper assessing planned social change impacts:**
 - “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor”
- **1997 Marilyn Strathern in paper regarding British University system audits:**
 - “When a measure becomes a target, it ceases to be a good measure”
 - Measure is often also called:
 - Key Performance Indicator (KPI)
 - Metric



“Cobra Effect”



Example	Measure:	Intended To:	Target:	Result:
Cobra Effect	Number of snake skins	↓ Snake population	↑ Snake skins	↑ Snake population



Education

• “Best College” Rankings

– Most statistically important metric for a university’s long-term financial viability

• Northeastern reverse engineered the statistical criteria, rose from 162 to 49 in 17 years and tripled tuition prices

- Capped classes at 19 students

- Easy online application

- Recruited heavily to decrease the ratio of students being accepted

- Encouraged lower-credentialed high school students to spend first semester abroad – excluded their GPAs from the GPA calculation

– Other schools admitted to cheating, lying about, or exaggerating the statistics

Example	Measure:	Intended To:	Target:	Result:
“Best College”	Ranking	↑ Information about Colleges	↑ Ranking	↑ Tuition for Top Schools



How do you reduce the impact?

- Use the scientific method to generate data
- Avoid the use of manipulated data
- Use data not generated by the organization being measured
- Utilize Measures of Effectiveness (MOEs) instead of Measures of Performance (MOPs)
- Design of the measure
 - Diversification
 - Secret metrics
 - Post-hoc specification
 - Randomization
 - Soft metrics
 - Limiting metrics
 - Abandoning measurement
- Considerations for the measure
 - Coherence
 - Structured Discussion / Compromise
 - Casual Forethought
 - Pre-Gaming
 - Monitoring Behaviors

“Metric design is an engineering problem, and good solutions involve both science and art.”



What is a MOE?

- **“MOEs are standards against which the capability of a solution to meet the needs of a problem may be judged. The standards are specific properties which any potential solution must exhibit to some extent. MOEs are independent of any solution and specify neither performance nor criteria.”**
- **MOEs are:**
 - **From the viewpoint of the stakeholder**
 - **“Stakeholders are therefore defined as those who...have demonstrated their need and willingness to be involved in seeking a solution.”**
 - **“Mission-oriented’ standards directed at the solution’s purpose rather than at the solution per se.”**
 - **“Presented as a statement and not a question”**
 - **Essential to the solution**
 - **Developed separately from the criteria**
 - **Able to be tested**
- **Characteristics of MOE indicators:**
 - **Can be measured quantitatively**
 - **Objective**
 - **Simple to state**
 - **Testable**
 - **Complete**
 - **Clear**
 - **States any time dependency**
 - **States any environmental conditions**
 - **Easy to measure**
 - **Structure so that they have measurable, collectible, and relevant indicators**
 - **Maximize clarity**



MOE vs MOP

- “An MOE refers to the effectiveness of a solution and is independent of any particular solution; a MOP refers to the actual performance of an entity.”

MOE	MOP	Indicator
Answers the question “Are we doing the right things?”	Answers the question, “Are we doing things right?”	Answers the question, “What is the status of this MOE or MOP?”
Measures purpose accomplishment	Measures task completion	Measures the data inputs to inform MOEs and MOPs
No hierarchical relationship to MOPs	No hierarchical relationship to MOEs	Subordinate to MOEs and MOPs
Often formally tracked in formal assessment plans	Often formally tracked in execution matrices	Often formally tracked in formal assessment plans
Typically challenging to choose the correct ones	Typically simple to choose the correct ones	Typically as challenging to choose as the supported MOE or MOP

J7 Commander’s Assessment Handbook (2011)

Relationship described using the analogy of effectiveness and efficiency:
“Effectiveness is how well something does its job. Efficiency is how well something does what it is doing. Therefore, efficiency can be high while effectiveness is low or even zero. Something can be done well even though it is the wrong job which is being done.”



Scoring Alternatives

- **Standards Based Grading**
 - Alternative to traditional grading that uses specific standards to assess both content and skills-based knowledge
- **Standards Based vs Traditional**
 - In traditional percentage system, failure is mathematically more likely to occur than any other grade
 - In a recent study of a course with 188 students, 16% of students had different grades under standards based (4% higher compared to traditional grading)
- **Key Principles**
 - Learning objectives and outcomes are explicitly stated and accessible
 - Multiple strategies to facilitate learning
 - Flexibility in the method to show mastery
 - Assessments are designed to test students on learning objectives
 - Course grades are based on students' ability to demonstrate understanding
 - Provide students with additional support
- **Four major types**
 - Contract Based
 - Mastery Based
 - Competency Based
 - Specifications Based



Contract Based Grading

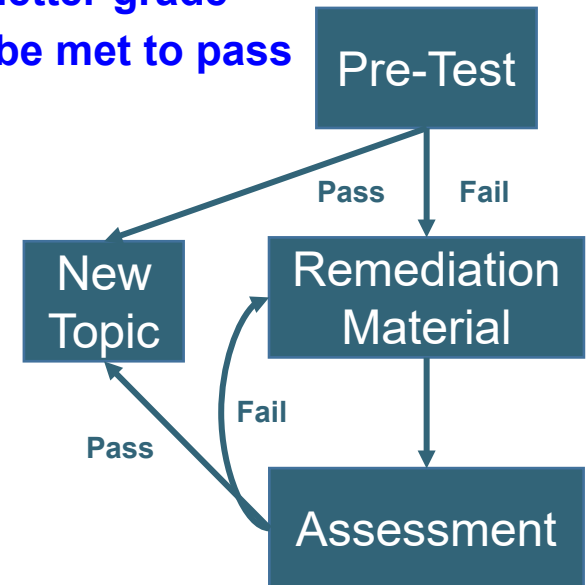
- **Contract between the student and the instructor on what must be completed and required mastery level**
 - **Student-created contracts or community-based contracts**
- **How to Implement**
 - **Develop learning objectives**
 - **Menu Items**
 - **Base contract to earn a B**
 - **Students begin work towards completion of contract**
 - **Instructor-generated feedback**
 - **Self-assessment**
 - **Peer review**
 - **Multiple student conferences with the instructor**

“Students also show shift from focusing solely on grades to being more involved with learning and over time develop a growth mindset.”



Mastery Based Grading

- **Mastery grading are provides students with learning objectives, allows students opportunities to show mastery of objectives and gives students multiple ways and attempts to master each objective**
- **How to Implement**
 - **Determine learning objectives**
 - **Define mastery**
 - **Establish grading**
 - **Require a specific set of standards to be met for each letter grade**
 - **Instructor determines which learning objectives must be met to pass**
 - **Incorporate flexibility**
 - **Open resubmission – resubmit as often as needed**
 - **Earned resubmission – resubmission meets a certain level**
 - **Token System – quantity during semester**
 - **Frequency or attempt limitations**





Competency Based Grading

- Incorporates aspects of Mastery grading while structuring learning into bundles or tiers that are associated with specific grades
- How to Implement
 - Create Learning Objectives
 - Critical/non-critical
 - Determine bundles/tiers
 - Place assignments and assessments into bundles or tiers
 - Assign bundles/tiers to grades
 - Implement resubmission or opportunity for re-test
 - Students determine grade they want for the unit or course by completing the bundle/tier of assignments

Table 1. Example of Competency-based Grading using Bundles for determining Student Grade within a specific unit. Students must earn all parts of a bundle to earn the grade associated with that grade bundle.

Grade	Unit Assignments	Unit Project	Unit Exams
Bundle 1 Requirements to earn a C in the course	Complete five of the assignments at 70% accuracy	Complete the project by addressing five of the components	Earn at least a 70% on each learning objective in the unit exams
Bundle 2 Requirements to earn a B in the course	Complete seven of the assignments at 80% accuracy	Complete the project by addressing eight of the components	Earn at least an 80% on each learning objective in the unit exams
Bundle 3 Requirements to earn an A in the course	Complete ten of the assignments at 80% accuracy	Complete the project by addressing all ten of the components	Earn at least a 90% on each learning objective in the unit exams

“Competency-based grading has been shown to increase student learning, engagement, and ownership of their learning”



Specifications Based Grading

DOC – Cost Estimating Directorate

- **Combination of Mastery Grading and Competency-Based Grading**
 - Specifications with pass/fail items
- **How to Implement**
 - Create student learning objectives
 - Develop assignments and assessments
 - Create bundle/tier for learning objectives

Table 1. Example of Specification Grading using Bundles for determining Student Grades. Students must earn all parts of a bundle to earn the grade associated with that grade bundle.

Grade	Weekly Quizzes	Midterm Project	Unit Exams	Additional Requirements
Bundle 1 Requirements to earn a C in the course	Earn at least a 70% on 10 of the 15 quizzes	Met 12 of the 20 specifications on the project rubric	Earn at least a 70% on each learning objective in the unit exams	No additional requirements
Bundle 2 Requirements to earn a B in the course	Earn at least an 80% on 12 of the 15 quizzes	Met 15 of the 20 specifications on the project rubric	Earn at least an 80% on each learning objective in the unit exams	Students complete one additional assignment for each unit with at least a 70% for each learning objective in the assignment

Table 2. Comparison of a traditional rubric and a specification rubric for writing conventions in a written assessment.

Traditional Rubric Criteria	Excellent (4 points)	Good (2 points)	Poor (0 points)
Writing Conventions	The paper is well written, has few misspelled words, uses proper grammar and punctuation, and follows MLA formatting.	The paper has many errors including misspelled words, improper use of grammar and punctuation and does not follow many aspects of MLA formatting	The paper is difficult to read because of many misspelled words, major issues with grammar and punctuation and does not follow any aspects of MLA formatting.

Specification Rubric Criteria	Satisfactory	Unsatisfactory
Writing Conventions Specifications		
Fewer than 5 words are misspelled in the paper.		
Only one or two sentences contain grammatical errors.		
All punctuation is correctly applied in the paper.		
MLA formatting is used correctly in 90% of the paper.		



New Scoring Method Ideas

- **Implementation of Specifications Grading**
 - GAO criteria as the “units” and bundle for the scores
 - Create specifications of GAO criteria utilizing Measures of Effectiveness (MOE)
 - Different bundling for MDA programs (existing, new, non-program estimates)
 - Community determined weightings
 - *Diversification*
 - *Structured discussion/compromise*
 - Create a resubmission process

GAO Criteria (Accurate)	Satisfactory	Unsatisfactory

MDA Program				
Score	Comprehensive	Well-Documented	Accurate	Credible
A				
B				
C				

- **Consider no scores**



Conclusions

- **Analysis of historical MDA scores are showing that the score is not measuring the quality of an estimate as intended**
- **MDA for 2024 is restructuring the reviews**
 - **Consideration for Goodhart's Law**
 - **Other grading mechanisms**
 - **Development of MOEs and indicators**
 - **No score and focus on actionable feedback**

GO FAST > THINK BIG





References

- **Building Less Flawed Metrics Dodging Goodhart and Campbell's Laws, David Manheim**
- **Building less-flawed metrics: Understanding and creating better measurement and incentive systems, David Manheim**
- **Goodhart's Law: Recognizing and Mitigation the Manipulation of Measures in Analysis, CNA**
- **“Goodhart’s law explains school decay”, American Institute for Economic Research, January 5, 2023, <https://www.aier.org/article/goodharts-law-explains-school-decay/>**
- **Coming to Grips with Measures of Effectiveness, Noel Sproles**
- **Establishing Measures of Effectiveness for Command and Control: A Systems Engineering Perspective, Noel Sproles**
- **“Alternative Grading for College Courses”, University of Nebraska-Lincoln <https://teaching.unl.edu/resources/alternative-grading/>**