32nd Annual INCOSE international symposium
hybrid event
Detroit, MI, USA
June 25 - 30, 2022

Multi-Disciplinary Insights into Measurement and Assessment for SoS

Jaci M Pratt
Defence Science and Technology Group
and The University of Adelaide
+61 0408 106 091
jaci.pratt@defence.gov.au

Stephen C Cook
The University of Adelaide
and Shoal Group Pty Ltd
+61 0418 829 946
stephen.cook@adelaide.edu.au

Abstract. The assessment of performance and quality is essential to the engineering and evolution of Systems of Systems (SoS); however, there has been limited discussion of SoS measurement and evaluation in the literature. This paper reviews the measurement and assessment literature across multiple disciplines to address this gap and applies constructive research to design principles and concepts for SoS evaluation. Relevant challenges and best practice elements are identified within each discipline for comparison against the nature and characteristics of Systems of Systems. These elements are then synthesized into three key areas necessary for SoS evaluation: measure development, determination of good metrics, and assessment. While further work has been identified and initiated on methods for the generation of such metrics, this study provides practitioners with the core practices and principles necessary to enable tailored and enhanced SoS assessment.

Introduction

The concept of Systems of Systems (SoS) arose in the late twentieth century through the increasing recognition that many large, often global, systems underpinning society arise through the integration of pre-existing systems. Often cited examples from the early literature include telecommunication networks, the internet, air traffic control, military forces and systems, transport systems, and international banking systems. The Australian Department of Defence (ADoD) has become interested in how Systems of Systems Engineering (SoSE) can aid in the integration of communications systems, command and control systems, and operational military capabilities into a joint force. A concise introduction to SoS and why they are of importance to all systems engineers can be found in the SoS Primer (INCOSE, 2018) and the INSIGHT Special Issue on SoS (INCOSE, 2016). In summary, a SoS is a "set of systems or system elements that interact to provide a unique capability that none of the constituent systems (CS) can accomplish on its own" (SEBoK, 2021). The primary taxonomy of SoS is derived from the degree of central governance which falls into one of four categories: directed, acknowledged, collaborative, and virtual; now standardized in ISO/IEC/IEEE 21841:2019. SoS are systems and accordingly, exhibit emergent behavior. As with all systems, there is a need for measures to characterize their emergent behavior as well as the progress being made towards achieving SoS performance and quality goals.

Unfortunately, there is limited literature available discussing measures for SoS and how they can be developed and used in evaluation and assessment. This study addresses this gap through the application of constructive research initially focusing on the state of measurement and supporting concepts / principles to address the SoS challenge. The next phase delves into identification of extant processes for measure development, the application of a selected set to case studies and the development of

SoS-specific measure generation processes and tools. Focusing on phase 1, this paper opens by reviewing the contributions on measurement from the systems and software engineering communities. Given that SoS of interest to the ADoD take the form of large, complex socio-technical systems, the study moves on to review the approaches across numerous disciplines that are used to perform the assessment and evaluation of such systems. The first of these is test and evaluation that originated from project-centric defense sector practices and focuses on the whether a project or service is achieving its design goals. This is followed by a review of approaches for the assessment and evaluation of larger-scale social systems, including force assessment, program evaluation, and enterprise improvement. Against this background, the paper examines contributions on the T&E of SoS and then progresses to synthesize best practice insights to inform SoS measurement, assessment, and evaluation practice.

Methodology

As part of broader studies on SoSE (Cook & Pratt, 2016 & 2020; Pratt & Cook, 2018), the authors have been working to improve the conceptual, theoretical, and practical basis for SoS measurement and assessment. An initial literature review of measurement and assessment directly tailored to SoS from across the systems engineering field identified limited coverage. This prompted the development of the wider, more comprehensive literature review and constructive research study.

SoS Engineering requires multi-disciplinary thought and action; hence, perspectives on evaluation and assessment from a variety of disciplines are needed. The study, therefore, selected and reviewed a series of socio-technical disciplines (Systems Engineering Validation, Verification, Test and Evaluation; Sociological Assessment; Training and Operational Assessment; and Enterprise, Program and Project Evaluation). From each of these disciplines, measurement and assessment challenges were identified, along with their associated perspectives on assessment and evaluation best practice. These practices were then compared against the fundamental characteristics of SoS to identify the most relevant and useful. These practices were finally synthesized to construct a series of best practice insights focused on three areas: SoS measure development, measures selection, and the assessment itself.

The Nature of Engineering Systems Measurement

Measurement can be defined informally using the statement below synthesized from the work of Finkelstein (1982) and Bush and Fenton (1990):

> *Measurement is the process of empirical, objective assignment of numbers or symbols to attributes of entities or events in the real world in such a way as to describe them.*

Technical measurement in Project Management (PM) and systems and software engineering is well established as shown through its inclusion as a technical management process in the top-level Systems Engineering (SE) standard, ISO/IEC/IEEE 15288:2015, and its own detailed standard: ISO/IEC/IEEE 15939:2007. The standards draw on valuable practice as documented in a series of INCOSE publications (Roedler and Jones, 2005; Miller et al., 2010; Carson et al., 2015). These documents treat measurement as a feedback control system, see Figure 1, where measurement informs action within SE and PM. Importantly, the literature makes clear that the value of measurement comes not from the act of measuring, but from the subsequent analysis and the action taken to correct variances from desired targets or to improve current performance. These fundamental concepts are equally valid for SoSE; however, the selection of measures and how they are used differs from conventional SE practice.

The INCOSE Measurement Primer (Miller et al., 2010) states that “measurement quantifies processes or work products with respect to the needs and objectives of the project or enterprise”. It also discusses the uses of measurement in terms of characterization, improvement, prediction, and evaluation, and provides a process for its planning and execution from the perspective of SE. Measures can be leading or lag indicators; that is, they can predict behavior, or provide a status or characterization of an activity (Rhodes et al., 2009). The combination of both lagging and leading indicators provides better decision making through a more thorough understanding of the current issues and gaps whilst gaining insight into expected impact of future system modifications.

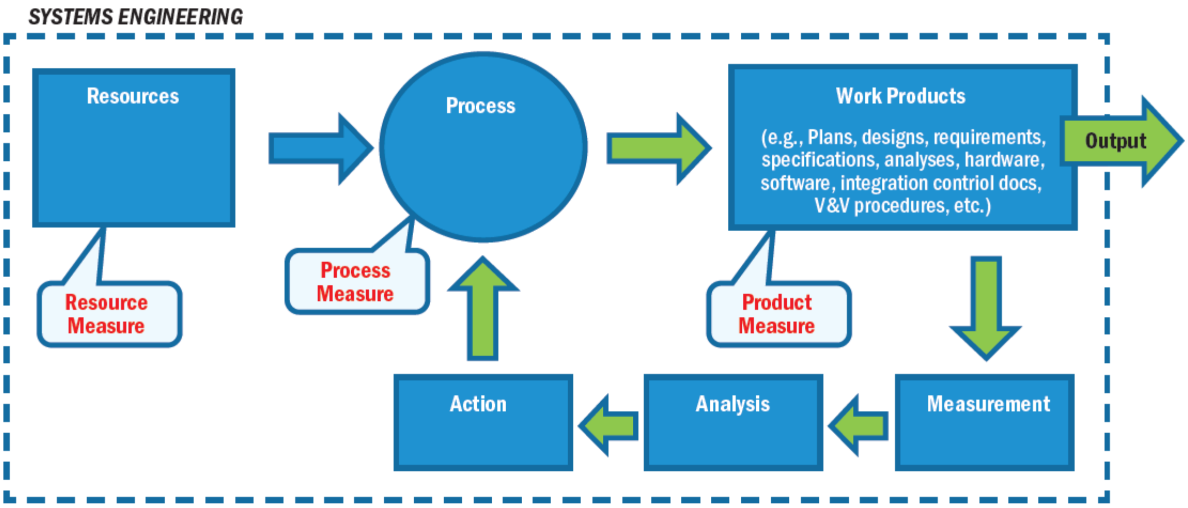

Figure 1. Measurement as a feedback control system (Miller et al., 2010)

The Motivation for Measurement

McGarry et al. (2001) and Carson et al. (2015), amongst many others, provide compelling reasons for measurement. These are synergized below and cast in a SoSE context. Measures are valuable to:

1. ***Track specific SoSE evolution objectives***. Tracking progress toward technical and management objectives enables better technical planning, resource allocation, and more informed decisions. Measurement helps identify areas of success and those needing improvement.
2. ***Identify and Correct Problems Early***. Measurement facilitates a proactive management strategy: fostering the early discovery, prioritization, and correction of problems more difficult and costly to resolve later.
3. ***Communicate effectively***. Measures provide the SoSE team with information on problems and status, the regular communication of which can increase awareness, reduce uncertainty and ambiguity, and improve organizational focus.
4. ***Allocate priority and focus on risk***. Measures help identify complexities, problems, and their root causes to permit active risk management. Concentrating on the most important measures improves direction and provides more time to focus on other things.
5. ***Inform and justify key decisions***. Measurement aids in tradeoff analysis, approach selection, identification of feasible solutions, and decision justification.
6. ***Assess quality***. Measures support engineering technical product and program quality.
7. ***Influence and shape SoSE evolution***. Measurement data and assessment insights assists the SoSE team to understand the SoS status and how to best move forward.

Types of measurement

It is useful to define the term *measure* before continuing. ISO/IEC/IEEE 15939:2007 defines the noun *measure* as a variable to which a value is assigned as the result of measurement and notes that the plural form *measures* is used to refer collectively to base measures, derived measures (a function of two or more values of base measure), and indicators (an evaluation of measured attributes derived from a model with respect to defined information needs). The reviewed references above identify three types of measures: resource, process, and product measures. **Resource measures** capture the expenditure of money and the utilization of human, computing, and infrastructure resources. **Process measures** relate to how well a given process or activity is working and they provide insights into process stability and improvement opportunities. Many generic process measures can be found in SEI (2010), such as cycle time and defect removal efficiency. These measures also provide useful information for estimating the resources needed for new projects. **Product measures** relate to the various work products. The principal product measures for major projects can be thought of as an interdependent hierarchy as described in Roedler and Jones (2005), comprising Measures of Effectiveness (MoE), Key Performance Parameters (KPP), Measures of Performance (MoP), and Technical Performance Measures (TPM). These are described below.

- ***Measures of Effectiveness (MoEs)*** are closely related to the achievement of the mission or operational objectives and are evaluated in the intended operational environment. MoEs are stated from the acquirer or customer viewpoint. They provide insight into how well the solution achieves its intended purpose.
- ***Measures of Performance (MoPs)*** characterize the physical or functional attributes relating to the system operation under specified testing and/or operating environmental conditions. Typically found in the system specification, they provide insight into system performance.
- ***Key Performance Parameters (KPPs)*** are a critical subset of the performance parameters representing the capabilities and characteristics that are so significant that failure to meet the performance threshold can be cause for the concept, project, or system being reassessed or terminated. Close monitoring of KPPs is considered a critical project success factor (Cook and Wilson, 2019).
- ***Technical Performance Measures (TPMs)*** focus on critical system technical parameters that if not achieved put the project at risk and are often derived from MoPs. They are used to assess design processes, compliance, and technical risks.

MoEs and MoPs are also common assessment measures used within Operational Analysis (OA). MoPs focus on the technical performance, or task actions, and are typically quantitative direct measurements. MoEs assess the impact of actions and their effectiveness in achieving objectives. They are usually more subjective (in both quantitative and qualitative form), and may be difficult to measure directly, often requiring the use of indicators or proxies. However, MoEs and MoPs do not scale up well to measure activities or organizations of greater scale and scope i.e. operational and strategic in military terms (Curtis and Bowley 1999). This reference proposed a hierarchy including higher-level measures, such as Measures of Outcome (MoO), Measures of Success (MoS), and Measures of Capability (MoC), to address these concerns.

Classes of measurement

Practical Software Measurement (PSM) defines seven common information categories (McGarry et al., 2001) for software intensive systems. The list below is drawn from that reference and modified to reflect a SoS context.

1. ***Schedule and Progress***: In normal projects, this addresses the achievement of project milestones and the completion of individual work packages. The SoSE team does not control CS projects but employs their outputs and capabilities. Application of SoS schedule and progress metrics will ensure achievement of roadmap milestones. Monitoring and aligning CS measures with these not only assures delivery of components vital to SoS capability increments but drives improved collaboration between CS and SoSE teams.
2. ***Resources and Cost***: This information category relates to the balance between the work to be performed and resources assigned in normal projects. As with the previous item, the SoSE

team will not commonly have control nor visibility of CS development; however, tracking SoS resource usage (where relevant) as well as monitoring the CS metrics supports achievement of goals. Where CS are encountering issues, the SoSE team may provide support to the CS (through resourcing or collaboration) to improve achievement of mutual goals.

3. ***Process Performance***: In normal projects, this information category relates to the capability of the supplier relative to project needs. As stated previously, in a SoS context, the SoSE team have little or no control over CS processes; however, metrics for the performance of the SoSE processes could be used.

4. ***SoS Performance and Behavior***: This category tracks the key performance and behavior measures of the SoS, such as transit times in a multi-modal metropolitan transport system.

5. ***SoS Evolution Rate***: This information category addresses the rate of change of the form and/or functional scope of the SoS and its CS.

6. ***SoS Quality***: This information category addresses the ability of the integrated SoS to support the user's needs without failure. Metrics for aspects such as reliability, sustainability, and other 'ilities', along with environmental considerations, would be included in this category.

7. ***Technology Effectiveness***: This information category addresses the viability of the proposed technical approach. It addresses engineering approaches such as use of commercial software components, software reuse, reliance on advanced software processes, and implementation of common software architectures. This may not be a high priority class for the SoSE team.

8. ***SoS Integration Assurance Measures***: This category includes metrics that are known to align with good SoS integration outcomes. Metrics in this category would be associated with: architectural design, dependency and interfaces, SoS risk analysis, SoS capability evaluation, SoS team management and competencies, road mapping and planning, stakeholder agreements, information management, and SoS culture (Cook and Unewisse, 2018).

9. ***Customer Satisfaction***: This information category addresses the degree to which products and services delivered by the SoS meet customer expectations. Indications of satisfaction may be obtained from customer feedback and the levels of customer support required.

Assessment and Evaluation

Assessment and evaluation are terms applied across many disciplines relating to the measurement of capability, with some using both terms and others just one; however, each have a slightly different focus. Evaluation provides a judgement of the system's capability, usually to a standard as set by the evaluator; it is product oriented. In contrast, assessment is an ongoing method of gathering information about system improvement; it is process oriented. Both of which are required for SoS.

SoSE processes include consideration of the performance and improvement of both the SoS and CS; however, the literature involving concepts and application of SoS assessment and evaluation is limited. To address this, as SoSE requires multi-disciplinary thought and action, perspectives of evaluation and assessment from a variety of disciplines are reviewed below, and an overview of their strengths and weaknesses for application to SoS extracted.

Test & Evaluation: the Systems Engineering perspective

Within the Systems Engineering community, evaluation is conducted in the context of either Validation & Verification (V&V) or Test & Evaluation (T&E). Verification focuses on confirming the achievement of specified requirements, whereas validation is an operationally focused process to derive "objective evidence that the system, when in use, fulfils its business or mission objectives and stakeholder requirements, achieving its intended use in its intended operational environment" (ISO/IEC/IEEE 24765:2010). V&V is often colloquially explained as "building it right" versus "building the right thing".

The United States Marine Corps Integrated T&E Handbook describes T&E as an iterative and incremental process occurring throughout the capability life cycle to "determine whether systems are operationally effective, suitable, survivable, and safe for intended use" (MCSC, 2010). T&E incorporates forms of both validation and verification through developmental, acceptance and operational T&E events. T&E and V&V assess system capability against the predetermined system requirements that describe specific user and system need. This means assessments are repeatable, structured activities that are typically quantitative with focus on technical aspects at a high level of detail. The system-of-interest and its context are pre-defined and influencing factors are managed to provide result certainty. This allows T&E and V&V to be conducted in the final phases of the project as the system components are progressively integrated (Walden et al., 2015).

In SoSE, where the SoS and CS are in continuous development, requirements are evolving, and execution occurs through asynchronous delivery of multiple projects at different lifecycle phases, evaluation aspects must be adjusted. T&E uses an evaluation framework of issues and measures, as described earlier (Roedler and Jones, 2005), as the process foundation to ensure effective assessment to decision makers. SoS measurement requires an even stronger focus on issues and measures hierarchies, given the natural structure and interdependencies within CS and the SoS. Furthermore, as SoSE follows a continuous iterative development model akin to contemporary software development, rather than a waterfall process, the issues, goals, and measures need to be revisited every iteration.

Assessment of social systems

Social systems are concerned with the network of relationships between individuals, groups, and institutions and their composition into a whole (Merriam-Webster, 2021). The conduct of interventions for improvement requires assessment; however, these human-based systems are difficult to measure due to their inherent variability and complexity.

Assessment of social systems requires the use of both qualitative and quantitative measures (Kaplan, 1964) from coding and classification, systemic rating and ranking scales, through to quantitative indices, and pure numerical data. There is significant pressure within the discipline (as in many other human sciences) for more quantitative analysis to align with the scientific process in search of increased validity. Cicourel (1964), however, warns that "imposing deterministic measurement systems on implicit theoretical concepts" is risky. Social and human data is open to contextual interpretation and perception (Cicourel, 1964), and any correlations or insights identified may simply be artifacts of quantification.

Babbie (2001) defines measurement as "careful, deliberate observations of the real world for the purpose of describing the objects and events" where observations may be direct, indirect or constructs. Measurement frameworks and structured thinking processes are used in social science to drive assessment from the core intent, situation, and functional purpose of the intervention (Kaplan, 1964). The measurement process moves through conceptualization, nominal definition, operational definition, to measures in the real world. Conceptualization is a key component, as it defines and compartmentalizes the constructs and indicators, connects observable data to constructs, identifies assumptions, and sets data collection (Babbie, 2001). This iterative process, covering the why, what, and how of measurement, both devises and applies the yardstick; thus, it is essential that the personal biases and judgements of the people involved must be acknowledged (Kaplan, 1964). Key measurement issues identified within social research include variables with multiple and ambiguous meanings, the impact of language choice on responses/data, reification of constructs, measurement impact on the system itself, variability in classification, comparison inability, and traceability/causality (Babbie, 2001; Cicourel, 1964; Kaplan, 1964).

Good research practice as identified with these references to mitigate SoS-relevant issues includes:

- Ensuring clarity in definition including range of variation and degree of precision (be specific and unambiguous) (Babbie, 2001).
- Taking multiple observations (indicators) and generate composite measures where the concept is open to interpretation (Babbie, 2001).
- Using an iterative methodology: identify, specify, and re-examine (concepts, models, definitions and data) (Babbie, 2001; Kaplan, 1964).
- Stating and reviewing assumptions, perceptions and interpretations and their impact continually (Babbie, 2001; Cicourel, 1964).
- Recognising the relationships implied through the assignment of numerals (i.e. order, additive and standard properties) and match analytical techniques (Kaplan, 1964; Babbie, 2001).

Assessing Joint Forces

Military forces across the globe conduct many forms of assessment for a variety of purposes. Such assessments provide assurance of performance and effectiveness, and span quantitative, qualitative, and combined methods. T&E of systems and sub-systems is used to provide confidence in technical capabilities and their performance. Training and certification assessments confirm competency against tasks and achievement of learning outcomes, and may be formative, summative, or diagnostic (ADoD, 2006). Such assessments are usually qualitative in nature and are performed by trained and independent evaluators. The scope and scale of military assessments can range from individual performance on single systems, to team assessments, through to coalition force effectiveness.

Assessing Joint Forces is a difficult yet essential task (Ochmanek, 2018). The complexity, scale, scope, uncertainty and dynamic nature of warfare, military objectives, capabilities, threats, and the environment (Ochmanek, 2018; Zvijac, 2012) create numerous challenges for joint force assessment. Despite this, military doctrine provides direction for evaluators to conduct assessments of plans, operations and effects (ADoD, 2006; OJCS, 2011). These assessments are usually Service-based and often focus on specific areas such as planning or operations. Currently, no systematic process exists within the US Department of Defense for joint force assessment (Marquis et al., 2016); however, some frameworks have been developed. The United States Agency for International Development Interagency Security Sector Assessment Framework (Chemonics International, 2010) provides such a process framework for the identification and prioritization of interagency security capability needs and gaps. Frameworks such as these have a conceptual basis similar to Checkland's Soft Systems Methodology (Checkland, 1999). Operations Analysis (OA) applies a similar framework approach through the development of an issue-measure hierarchy for analysis and assessment of military socio-technical systems, such as the MOP-MOE-MOC-MOO hierarchy (Curtis & Bowley 1999).

Many issues have been raised with the development and conduct of joint force assessments (Marquis et al., 2016; Meiter, 2012; Zvijac, 2012):

- Underpinning *concepts and definitions* are poorly developed, expect defined end-states, lack connection to the assessment rationale, and do not consider complexity (i.e. ambiguity, unpredictability, and change).
- *Processes* are convoluted, poorly resourced, not comprehensive, lack alignment with civilian best practice, and are inconsistent across the domain/area.
- Assessment *teams* are typically understaffed and undertrained (particularly formally).
- Measures or assessment *indicators*, although often abundant, are poorly crafted, difficult to collect and ambiguous, providing minimal rigour or evidence for later analysis and decision-making.

Notwithstanding, there is discourse regarding what constitutes good practice. Key success factors for joint force assessment identified within the literature include:

- *Concepts* must be based on an agile framework and sound methodology that inherently accepts uncertainty, incompleteness, and bias. Evidence is gathered from multiple perspectives and methods to seek steady improvements rather than attaining a specified end-state (Welshans et al., 2016; Zvijac, 2012).
- Assessment *processes* should be clear, consistent, holistic, systematic, and evolve. They will require well-coordinated reporting and tools, and teams and methods matched to the process purpose (Marquis et al., 2016; Welshans et al., 2016; Zvijac). Assessments should seem "less like testing and more like discovery" (Welshans et al., 2016).
- Evaluation *teams* need to be well-trained, independent, and dedicated to the assessment function (Marquis et al., 2016; Ochmanek, 2018).
- Use of a limited set of consistently applied strategic / aggregate *indicators* ("less is more") (Marquis et al., 2016; Welshans).

Enterprise, Program and Project Evaluation

In Enterprise Systems Engineering (ESE), an inherently qualitative approach is taken, where the terms *evaluation* and *assessment* are used to determine if the "enterprise is heading in the right direction" (Martin, 2010). The enterprise architecture, which is directly connected to the business goals, is used in the development of enterprise success measures. Rather than utilizing the standard systems engineering T&E methodology (comparison against requirements), Martin proffers the use of multi-methodological break point analysis (Martin, 2010) that employs modelling and simulation at multiple scales conducted in a continuous, iterative fashion.

Program Evaluation (PE) is a methodology to evaluate whether (often very large) programs are producing their intended effects. Owen (1993) identifies five program classes commonly evaluated using this methodology: educational programs, advisory programs (e.g. health, education), regulatory programs (e.g. reducing drink driving), and products or services (such as public transport, provision of internet services, etc. typically seen as SoS). PE uses a helical cycle of enquiry, starting with a reflection phase that considers discrepancies between the observed and the expected (e.g. declining use of public transport despite a program to increase participation). In the subsequent design phase, stakeholders are identified, and the evaluation is planned based on a consensus of the needs. The fieldwork stage requires evaluators to become anthropologists: observing and interacting with the system and its participants to develop understanding. Evaluators must be skeptical and challenge the established norms to achieve a rich understanding of the situation that is refined through feedback with the evaluated group. These evaluation findings inform the final planning phase that produces socially inclusive options for change.

While PE is often considered to follow the interpretive tradition of developing an understanding of perspectives, experiences, and expectation of stakeholders, it can also follow the positivist approach that seeks to use objective, observable and essentially quantitative evidence, and the critical-emancipatory tradition that acknowledges that evaluation takes place in socio-political contexts. This methodology is well suited to evaluating socio-technical, enduring SoS like public transport, health care, and defense that are continually evolving in the face of changing social norms. The measures in PE are crafted bespoke to meet the needs of each evaluation but there is a rich literature from which to glean commonly used measures for particular classes of systems.

The Performance Management Process (PuMP), developed by Barr (2014) is a well-documented and structured methodology for designing performance measurement for organizations and projects. It is a practical and pragmatic process that focuses on goals, clarity of terms, and recognition of the people component. A key aspect of the approach is the development of a performance culture. The basis of the process is overcoming eight "bad habits" (Barr, 2017) in mainstream measurement practice: 1) measures shouldn't judge people; 2) ambiguous goals; 3) brainstorming measures; 4) not involving stakeholders; 5) rushing to tools, visualizations, and reports; 6) making conclusions on limited data; 7) using complex reports and too many measures; and 8) treating symptoms not problems. The PuMP process takes teams through eight steps to measure and achieve organizational goals (Barr, 2014): 1) understanding the purpose of measurement; 2) identifying measurable results; 3) designing meaningful measures; 4) encouraging measures buy-in; 5) implementing measures; 6) interpreting insights; 7) reporting performance; and 8) determining and achieving targets.

SoS Evaluation

System-of-Systems Engineering (SoSE) is a relatively new field that requires additional considerations in its conduct to that of traditional product-based Systems Engineering (SE). SoSE is not just a technical process; it must cover the socio-technical and capability element aspects. Indeed, SE itself has acknowledged the requirement for more socially aware terminology and processes (e.g. business/mission analysis within the INCOSE SE Handbook (Walden et al., 2015)). Engineering methods have been developed for specific problem areas, types, or scales; however, there is no "single formal methodology" (Guariniello et al., 2019) in place, with most explicitly developed or "hand-crafted" (Wanderley et al., 2018). This is due to a greater need for tailoring given the spectrum of SoS characteristics and types, and their impact on actions. Many of these specialized methodologies are focused on design, architecting, or developmental phases and are often conceptual in nature.

Most literature on the assessment and evaluation of SoS has originated from the systems engineering (SE) community in the defense domain. It indicates that SoS T&E must occur in different ways, for different purposes, and throughout the SoS lifecycle such as the US DoD wave model (Dahmann et al., 2010; Dahmann, 2012; Hess & Valerdi 2010; Meyers & Hester, 2012; Neves et al., 2018). Differences due to the fundamental nature of SoS have been investigated and challenges identified (Neves et al., 2018). Proposals for addressing these challenges include taking lessons from other disciplines such as software engineering (Neves et al., 2018) and adaptation of the SE processes to account for SoS (Browning & Honour, 2007; Dahmann, 2012; Dahmann et al., 2010; Honour & Browning, 2005; Herdlick, 2011). The SoSE process is seen as the "starting framework to address SoS T&E challenges" (Dahmann, 2012) thereby intrinsically linked and dependent on the successful implementation of SoSE. Within the US DoD SoSE wave model process (Dahmann, 2012), a conceptual process and artefact suite focus on addressing risk for acknowledged SoS. While this provides a basis for T&E, it "leaves open the question of how to assess the end-to-end behavior of the SoS in context" (Dahmann 2012). A more holistic view measures the life-cycle value of enduring systems using key parameters in a stakeholder-driven process (Honour and Browning 2005; Browning & Honour 2007). This acknowledges the qualitative and subjective aspects of value measurement and stakeholder perception, the evolution of these over time, and the combination of measures into holistic metrics.

While challenges are discussed in the references, the basics of good practice in SoS T&E are also offered (Browning & Honour, 2007; Dahmann et al., 2010; Dahmann, 2012; Honour & Browning, 2005; Herdlick, 2011; Meyers & Hester, 2012; Neves et al., 2018) and are summarized here as:

- Use stakeholder-driven continuous-improvement processes focused on risk areas
- Base T&E on a thorough understanding and application of SoS context and need
- Acknowledge evolution of metrics, stakeholders, and their perception through incremental and continuous T&E through the evolution of the SoS
- Employ evidence-based assessment across multiple sources and time
- Set SoS objectives and metrics at both SoS and CS level
- Use mixed metrics – quantitative, qualitative and constructs; functional and non-functional

Insights toward best practice for SoS

Assessment of SoS is unlike any single-discipline-based assessment method due to the fundamental nature and characteristics of SoS. These include independence (operational and managerial), geographic distribution, emergence, evolutionary development, decentralization, networked, and heterogeneous (Sage & Cuppan 2001, Gorod et al 2008). Informed by the multi-disciplinary reviews above and an understanding of these fundamental characteristics of SoS, a series of draft practice insights have been synthesized for SoSE. The insights have been focused into three areas: developing measures, the evaluation of measures (i.e. choosing good measures), and assessment conduct.

Best practice for developing SoS measures

This review discovered that each discipline had experience aligned to SoS characteristics, with many issues and practices identified as relevant for SoSE. In earlier work on software metrics, Cook et al. (2021), based on Jones et al (2020), identified principles seen as valuable for practitioners of large system and software developments. These have been synthesized with the insights found in the literature reviewed to provide best practice principles for developing SoS measures.

1. ***Metrics are evidence and must be driven by purpose***: Measures are a decision tool relating to stakeholder views of SoS value. They should be traceable to agreed SoS goals and strategies and provide insights that drive timely, corrective action. Measures provide the most relevant and feasible evidence of achievement of SoS goals.
2. ***Context guides selection***: Metrics (especially higher-level metrics) need to be tailored to the context i.e. SoS type, development approach, environment, etc.
3. ***Measures must be agreed by stakeholders and have identified owners***: Stakeholders need to connect measures to their information extraction and decision-making needs. These measures must provide perceived benefit to CS and SoS stakeholders and will drive collaborative behaviors essential to SoS success. A stakeholder must be identified to champion data capture, analysis and reporting of each measure.
4. ***Clarity is all***: Measures need to be clear, understood, unambiguous, well-defined, and specific This includes both what is being measuring and how it is measured.
5. ***Measures must be mixed, acknowledging the value of qualitative & subjective data***: SoSE require the use of both qualitative and quantitative, functional, and non-functional measures. Understanding SoS requires knowledge of cultural, perceptual, and holistic information, notably qualitative (from interviews, surveys, etc). Enumeration of this data is common and useful; however, artifacts of quantification must be carefully considered.
6. ***Metric Parsimony***: A minimum practical set of measures should be selected and connected in hierarchy. The top-level SoS measures should be limited, indicating current priorities and focus areas; in fact, Sillitto (2014) suggests that 10 would be too many.
7. ***Automate and reuse where possible***: Noting principles 1 and 6, use automated or extant data and collection processes where possible. Utilizing CS measures to support understanding of SoS capability is important.

Characterizing Good SoS Measures

Following the practices described above will assist in generating a list of measures and metrics for SoS; however, choosing the most appropriate measures for the SoS and its context to meet practice #6 (Metric Parsimony) is a difficult step; more so, given the natural tension between the characteristics or quality attributes as described above. Considerable work has been done in characterizing good measures across the disciplines discussed in this paper. Operations Research which uses a hierarchy of measures has identified the following qualities of good measures: meaningful, endorsable, quantifiable, purpose-driven, useful, and timely (Curtis and Bowley, 1999). The PuMP process (Barr, 2017) utilizes two characteristics to drive measure selection: strength (correlation with system goals) and feasibility (ease with which the measurement can be performed). Key characteristics of good social measures (Babbie, 2001) include precision, accuracy, reliability, and validity (where validity covers face, criterion-related, construct and content validity). ISO standards, SE measurement guidebooks, and the wider measurement literature (e.g. Sydenham, 1982) also provide guidance on selection of the best metrics within SE and the technical engineering disciplines.

The art of measure selection is choosing the characteristics most relevant based on the context. The natural tensions and alignments between these measure characteristics needs to be understood and exploited. SoSE teams should choose a limited set of the most suitable characteristics based on their relevance and applicability to the SoS context. These characteristics can then be applied to ensure both the individual measures and complete set meet current needs.

Characteristics of good measures discovered within the reviews have been revised for applicability to SoS and amalgamated to generate guidance for SoS metric evaluation. When inspecting the characteristics of measures discussed across the disciplines, it was noted that some characteristics apply directly to individual measures while others apply to the set of measures; hence this guidance is provided in two sections.

Characteristics of Good Individual SoS Measures. The list below, synthesized from information gathered in the review (Babbie, 2001; Barr, 2014; Curtis & Bowley, 1999; Roedler & Jones, 2005), provides characteristics suited to the SoS problem that can be used to assess the quality and relevance of individual measures. The SoSE team could choose a small set of these characteristics, aligned with purpose and priorities, to maintain process efficiency.

1. ***Relevance / Strength***: Measures should be clearly connected to purpose and easily relatable for clients and analysts, such that they are easily endorsed. Choose the measures that are most highly correlated with SoS goal achievement.
2. ***Simplicity and Feasibility***: Measures (indicators) should be logical and easy to understand. It should be clear how the measure is collected and analyzed.
3. ***Cost effectiveness***: Measurement costs money. It is important to maximize the return on investment based on the value provided and the cost to collect the data. Re-use must be traded off with alignment of the measurement to purpose.
4. ***Accuracy and Precision***: Accuracy is the extent to which the value of a measure truly represents the entity being measured. The accuracy of the measurement needs to be sufficient to provide the information required. Precision relates to ability to discriminate between the attributes of measurement.
5. ***Validity:*** The extent to which the measurement adequately represents reality. Typical types of validity include face-, predictive-, construct- and content-validity. There is a natural tension between validity and accuracy/repeatability.
6. ***Repeatability***: This characteristic focuses on the ability of a measure to give the same value of the attribute each time it is measured and is important for comparing measures. It should be noted that repeatability is different from accuracy i.e. systematic measurement errors can result in low accuracy but good repeatability.
7. ***Predictability and boundedness***: Each measure should have time-phased threshold values and tolerance bands that can be predicted and substantiated during SoS realization.
8. ***Timeliness***: The measure must provide the needed information within time and frequency required to support corrective action.

Characteristics of a Good Set of Measures. Some quality attributes for metrics are only fulfilled through the total set of measures chosen. The characteristics of a good set of measures identified and combined from the review are discussed below. These can be used to inform both measurement design and confirmation of the final set of measures.

1. ***Saliency***: The set of measures should represent the most significant qualifiers of the SoS with each measure providing insight into key performance parameters and mission success.
2. ***Completeness and cohesiveness***: The set of measures should cover the most important classes of measurement, and as a minimum, performance, quality, and resources. The set of SoS measures should represent the best balance of needs across all stakeholders to win stakeholder buy-in and drive the SoS towards agreed goals.
3. ***Cost effectiveness***: The measures set should collect the information of most value for the nominated measurement budget.
4. ***Balance across quality and speed***: Balance the measures of delivery and quality across the metric set to ensure a best value solution based on SoS objectives.
5. ***Minimal disturbance***: The set of measures and their conduct should cause minimal disturbance to the functionality and capability of the SoS.

SoS Assessment Best Practice

This section provides a summary of the initial insights on the interpretation of information and process of assessment to support SoSE decision-making. The insights have been compiled and are discussed in their clusters.

1. ***Goals are the primary focus***: SoS value and outcomes should be agreed up-front, noting SoS goals will change over time. There are many ways to achieve the same goal in a SoS – via the CS, combinations of CS, the SoS itself, and across all elements of capability.
2. ***Incremental and iterative process***: SoS Assessment is not a question-answer process nor a single event; it is never complete just as the SoS is never "finished". Rather it is a continuous, incremental, and iterative process that edges the SoS ever forward through improvement. Measures should be collected on an ongoing basis from tests, analyses, or models, allowing prediction of, and insights into, future delivered SoS performance.
3. ***Evolution and emergence impact assessment scope and focus***: SoS assessment must adapt throughout SoS evolution. Focus points are determined based on risk / priority areas with review points scheduled and aligned to SoS milestones.
4. ***Stakeholders must be involved throughout***: The complete breadth of stakeholders (CS and SoS) must be involved in identifying and agreeing objectives, priorities, risks, and metrics. The SoSE team must recognise the alignment and tension between CS and SoS strategic visions and use metrics appropriately to seek convergence where possible.
5. ***SoS assessment is socio-technical***: Technical assessments are insufficient; non-functional and human aspects have greater impacts due to the strong value and user basis of SoS.
6. ***Metrics and Measures Hierarchy:*** Development of an agile framework of interdependent measures and indicators (proxies), across the CS and SoS, can provide incentives and drive behaviour (3^{rd} order effects of emergence and adaptation). Both leading and lagging indicators must be addressed to enable plan-and-adjust SoSE practices. Metrics will be different for different SoS types, lifecycle, status, etc.
7. ***Evidence is needed from multiple sources and methods***: Multi-methodological evidence, that is taken over time, at different points, and from multiple perspectives, provides greater ability to interpret the information and generates triangulation of results.

8. ***SoS metrics require interpretation***: The results of assessment of SoS suggest points of interest that demand investigation and require interpretation within the context and knowledge of the measurement schema. As SoS assessment is about improvement, iteration and sufficiency, trends and changes are important, not absolute values. “Rough baselines” can assist.

Conclusions and Further Work

This paper has described the outcomes of phase 1 of a constructive research project investigating measurement for SoS. It has provided an overview of different conceptual approaches to measurement and assessment across several disciplines relevant to SoS and extracted pain points and best practice elements from these. A synthesis of these lessons into guidance on measure development, evaluation of metrics and assessment best practice for SoS has been developed and presented.

SoSE seeks beneficial outcomes to all stakeholders (CS and SoS) through the improvement of the effectiveness and efficiency of the holistic SoS and its parts. This requires clarity of agreed needs and collaborative behaviors to achieve them; both of which can be driven using good measures. The concepts and principles identified here for SoS measurement, start this practice. However, measurement is one aspect only and can be over-valued; SoS metrics are simply indicators, and reliability is an issue for continuously evolving SoS in an uncertain environment. Kaplan (1964) reminds us that “our ability to measure depends on our conceptualization and knowledge of the situation, and the skill to the process of measurement”. Reliability of metrics can be improved through good understanding, good skills and good processes for measure derivation and conduct.

While the best practice concepts and principles presented here are essential for the development and application of SoS assessment, they do not provide clarity in methodology for the generation of SoS metrics and assessment processes. The second phase of this research focuses on the methodology and processes for SoS metrics generation to enable complete, coherent, and systematic SoS assessment practice. Through further reviews, design and case studies, this phase will deliver the supportive processes needed to ensure SoS measurement and assessment is both achievable and practical.

References

Australian Department of Defence (ADoD) 2006, *The Defence Training Model*, Doctrine and Training Series, ADFP 7.0.2, Australian Defence Headquarters, Canberra, ACT.

Babbie, R 2001, *The Practice of Social Research*, Wadsworth Thomson Learning

Barr, S 2014, *Practical performance measurement: Using the PuMP blueprint for fast, easy and engaging KPIs*, Samford: PuMP Press.

Barr, S 2017, *Prove it!: How to create a high-performance culture and measurable success*, John Wiley & Sons.

Browning, TR & Honour, EC 2005, "Measuring the Lifecycle Value of a System", Proceedings of the INCOSE International Symposium, 15(1), pp. 1530-45.

Bush, ME and Fenton, NE 1990, "Software measurement: a conceptual framework", *Journal of Systems and Software*, 12 (3), pp. 223-231.

Carson, RS, Frenz, PJ & O'Donnell, E 2015, *Project Manager's Guide to Systems Engineering Measurement*, INCOSE-TP-2015-001-01, INCOSE.

Checkland, P 1999, *Systems Thinking, Systems Practice (including Soft Systems Methodology: A 30 year retrospective)*, Chichester: John Wiley & Sons Ltd.

Chemonics International 2010, *Interagency Security Sector Assessment Framework*, Washington, D.C.: United States Agency for International Development, October 1, 2010.

Cicourel, AV 1964, *Method and measurement in sociology*, Oxford, England: Free Press of Glencoe.

Cook, SC & Pratt, JM 2016, "Typology dimensions for classifying SoSE problem spaces", *11th System of Systems Engineering Conference (SoSE)*, 12-16 June 2016.

———— 2020, Advances in systems of systems engineering foundations and methodologies, *Australian Journal of Multi-Disciplinary Engineering*, 17 (1), pp. 9-22.

Cook, SC & Unewisse, MH 2018, "An Austere SoS Integration Assurance Methodology", In *INCOSE International Symposium*, 28(1), pp 1478-1493.

Cook SC & Wilson SA 2019, "The Enduring Path to System Success: Investment in Quality Early-phase Systems Engineering", *INCOSE International Symposium*, 29(1), pp. 817-835.

Cook, SC, Samalam, A and Unewisse, MH 2021, "A Metrics Framework to Facilitate Integration of Disaggregated Software Developments", In *INCOSE International Symposium*, 31 (1), pp 1193-1210.

Curtis, N & Bowley, DK 1999, "Hierarchical systems of enquiry for analysis of the Land Force", In: *Proceedings of the Australian Society for Operations Research (ASOR)*, Queensland.

Dahmann, JS, Lane, JA, Rebovich, G & Lowry, R 2010, "Systems of systems test and evaluation challenges", *5th International Conference on System of Systems Engineering (SoSE)*, 22-24 June.

Dahmann, JS 2012, "Integrating systems engineering and test & evaluation in system of systems development", *2012 6th Annual IEEE Systems Conference (SysCon)*, 19-22 March 2012, Piscataway, NJ, USA.

Finkelstein, L 1982, "Theory and philosophy of measurement", In *Handbook of measurement science*, 1, pp. 1-30.

Gorod, A, Sauser, B & Boardman, J 2008 "System-of-Systems Engineering Management: A Review of Modern History and a Path Forward", *IEEE Systems Journal*, 2 (4), pp. 484-99.

Guariniello, C, Raz, AK, Mockus, L & DeLaurentis, D 2019, "Experiences in Evolving System-of-Systems Engineering Methodology to Address Pain Points", *14th Annual Conference System of Systems Engineering (SoSE)*, 19-22 May 2019.

Herdlick, B 2011, "Establishing an operational context for early system-of-systems engineering activities", *Systems Research Forum*, 5 (2), pp. 109-141, World Scientific.

Hess, JT & Valerdi, R 2010, "Test and evaluation of a SoS using a prescriptive and adaptive testing framework", In *5th International Conference on System of Systems Engineering (SoSE)*, 22-24 June 2010, pp.1-6.

Honour, EC & Browning, TR 2007, "Dynamic optimization of systems of systems using value measurement", *Journal of Integrated Design and Process Science*, 11(2), pp. 33-53.
INCOSE, 2018, *INCOSE Systems of Systems Primer*, INCOSE-TP-2018-003-01.0, INCOSE.
INCOSE 2016, Special Edition on Systems of Systems, *INSIGHT*, 19 (3).
ISO/IEC/IEEE 15288:2015, Systems and software engineering – System life cycle processes, ISO, IEC, IEEE.
ISO/IEC/IEEE 15939:2007, Systems and software engineering – Measurement process, ISO, IEC, IEEE.
ISO/IEC/IEEE 21841:2019, Systems and software engineering – Taxonomy of systems of systems, ISO, IEC, IEEE.
ISO/IEC/IEEE 24765:2021, *Systems and software engineering – Vocabulary,* ISO, IEC, IEEE.
Jones, CL, Golanz, B, Draper, GT & Janusz, P 2020, *Practical Software and Systems Measurement Continuous Iterative Development Measurement Framework*, Version 1.05 15 June 2020, PSM, NDIA & INCOSE.
Kaplan, A 1964. *The conduct of inquiry*. San Francisco: Chandler.
Marine Corps Systems Command (MCSC) 2010, *US Marine Corps (USMC) Integrated Test and Evaluation Handbook*, 6 May 2010.
Marquis, JP, McNerney, MJ, Zimmerman, SR, Archer, M, Boback, J & Stebbins D 2016, *Developing an assessment, monitoring, and evaluation framework for US Department of Defense security cooperation,* RAND Corporation-National Defense Research Institute, Santa Monica.
Martin, JN 2010, "An Enterprise Systems Engineering Framework", 20th Anniversary International Council on Systems Engineering (INCOSE) International Symposium.
McGarry, J, Card, D, Jones, C, Layman, B, Clark, E, Dean, J & Hall, F 2001, *Practical Software Measurement: Objective Information for Decision Makers*, Adison Wesley.
Meiter, JS 2014, "Assessment: A Military Methodology in Need of an Overhaul", MSc, Joint Advanced Warfighting School, Joint Forces Staff College, National Defense University.
Merriam-Webster.com online dictionary, s.v. "social system," accessed October 28, 2021, https://www.merriam-webster.com/dictionary/social%20system.
Meyers, TJ & Hester, PT 2012, "Measuring system of systems performance", *International Journal of System of Systems Engineering*, 3, pp. 277-289.
Miller, C, Carson, RS, Fowler, DJ, Gantzler, DJ & Roedler G 2010, *INCOSE Systems Engineering Measurement Primer v2.0,* INCOSE-TP-2010-005-02, INCOSE.
Neves VdO, Bertolino A, De Angelis G & Garcés, L 2018, "Do we need new strategies for testing systems-of-systems?", *Proceedings of the 6th International Workshop on Software Engineering for Systems-of-Systems*, Gothenburg, Sweden: Association for Computing Machinery, pp. 29–32.
Ochmanek, DA 2018, *Improving Force Development Within the U.S. Department of Defense: Diagnosis and Potential Prescriptions,* RAND Corporation (Santa Monica, California). https://www.rand.org/pubs/perspectives/PE302.html.
Office of the Joint Chiefs of Staff (OJCS) 2011, *Commander's Handbook for Assessment Planning and Execution, Joint and Coalition Warfighting*, Joint Staff J7, OJCS ed., Suffolk, Virginia.
Owen, JM 1993, Program Evaluation Forms and Approaches, Allen & Unwin, ISBN 1 86373 462 7.
Pratt, JM & Cook SC 2018, "A principles framework to inform Defence SoSE Methodologies", In *Disciplinary convergence in systems engineering research*, 197-213. Springer.
Rhodes, DH, Valerdi, R & Roedler, GJ 2009, "Systems engineering leading indicators for assessing program and technical effectiveness". *Systems Engineering*, 12 (1), pp. 21-35.
Roedler, GJ & Jones, C 2005, *Technical Measurement. A Collaborative Project of PSM, INCOSE and Industry* (INCOSE-TP-2003-020-01). PSM and INCOSE.

Sage, AP & Cuppan, CD 2001, “On the Systems Engineering and Management of Systems of Systems and Federations of Systems”, *Journal of Information, Knowledge, Systems Management*. 2, pp. 325–45.

SEBoK, 2021. *Systems of Systems Knowledge Area*, viewed 10 November 2021, https://www.sebokwiki.org/wiki/Systems_of_Systems_(SoS).

Sillitto, H 2014, *Architecting systems: Concepts, principles and practice*, College publications.

Sydenham, PH 1982, *Handbook of Measurement Science Volume 1 Theoretical Fundamentals*, Wiley.

Systems Engineering Institute (SEI) 2010, CMMI® Measurement and Quantitative Management Process Areas, Version 1.3, Software Engineering Institute.

Walden, DD, Roedler, GJ, Forsberg, KJ, Hamelin, RD & Shortell, TM 2015, *Systems Engineering Handbook*, INCOSE-TP-2003-002-04, Wiley.

Wanderley, GMP, Abel, M, Paraiso, EC & Barthès, JA 2018, “MBA: A Framework for Building Systems of Systems”, *13th Annual Conference on System of Systems Engineering (SoSE)*, 19-22 June 2018.

Welshans, JS, Owen, C, Mulvehill, AM, Hickey, CW & Farrell, RJ Jr. 2016, “Operational assessment: so how are we doing?” *Air & Space Power Journal*, 30 (4), p 18.

Zvijac, DJ 2012, *An Alternative Approach for Operational Assessment*, Report #: DRM-2012-U-002260-Final, CNA, cna.org.

Biographies

Jaci Pratt is a System-of-Systems Engineering and Integration researcher in the Defence Science and Technology (DST) Group within the Australian Department of Defence. She has worked in DST Group for over 30 years in areas as diverse as Modelling & Simulation, RF Stimulation, Experimentation, IT, Systems Analysis, and Test & Evaluation supporting Land, Air, Maritime and Intelligence domains. Jaci holds a BSc (Hons) from Flinders University and is completing a PhD in System-of-Systems Engineering with the University of Adelaide.

Stephen Cook is a Systems Engineering Advisor with Shoal Group Pty Ltd. He is also the Professor of Defence Systems at the University of Adelaide where he researches and teaches system engineering and complex project management. Until June 2014 he was the Professor of Systems Engineering at the University of South Australia where he led numerous research concentrations for over 15 years. Preceding this he accumulated 20 years of industrial R&D and SE experience spanning aerospace and defence communications systems. Prof Cook, PhD, is an INCOSE Fellow, a Fellow of Engineers Australia, and a Member of the Omega Alpha Association.