REVIEW OF MEASUREMENT PROPERTIES FOR THE USE OF INTEGRATED SYSTEM VALIDATIONS

Jinkyun Park[1], Jung Hwan Kim[2]

[1] 989-111 Daedeok-daero, Yuseong-gu, Daejeon, 34057, Republic of Korea, kshpjk@kaeri.re.kr
[2] 989-111 Daedeok-daero, Yuseong-gu, Daejeon, 34057, Republic of Korea, jhkim21@kaeri.re.kr

EXTENDED ABSTRACT

One of the important approaches ensuring the safety of a nuclear power plant (NPP) is to carry out diverse verification and validation activities including integrated system validation (ISV). The underlying idea of the ISV is to make sure that the performance of human operators who have to accomplish safety-critical tasks can be maintained without any degradation during the operation of NPPs. This implies that the development of ISV scenarios that contain representative safety-critical tasks is very important. In addition, it is also crucial to select appropriate measurements that can properly capture the change of human performance observed from ISV scenarios. In this regard, regulatory guidelines of the ISV suggest several dimensions that should be used as criteria for the selection of human performance measurements (i.e., performance measurement criteria). Typical dimensions include (1) Diagnosticity, (2) Sensitivity, (3) Construct validity, (4) Intrusiveness, (5) Reliability, (6) Objectivity, (7) Simplicity, (8) Impartiality, and (9) Resolution [1-3]. Table 1 shows the definition of each property available from existing literature.

TABLE I. Definitions of Nine Properties

ID	Property	Definition
1	Diagnosticity	Measurement should be able to identify the cause of unacceptable human performance.
2	Sensitivity	Measurement should detect the changes in the concept it measures
3	Construct Validity	Measurement should represent what it claims in terms of human performance.
4	Intrusiveness	Measurement should not intervene participants in terms of physical and psychological aspects.
5	Reliability	Human performance should be similar when it is repeatedly measured in an identical condition.
6	Objectivity	Measurement should provide objective information instead of subjective information.
7	Simplicity	Straightforward measurement should be used to make sure of its applicability.
8	Impartiality	Measurement should be equally capable of reflecting good as well as bad performance.
9	Resolution	Measurement should reflect an appropriate level of detail to permit meaningful analysis.

From Table I, it should be noted that the eighth and ninth property ('Impartiality' and 'Resolution') seem to largely overlap with those of other dimensions. For example, the definition of 'Impartiality' is that "Measurement should be equally capable of reflecting good as well as bad performance." With this definition, it is reasonable to expect that a specific human performance measurement will satisfy impartiality if it can provide a reliable, diagnostic and objective value that is straightforward to understand without any additional knowledge and/or information. Similarly, since 'Resolution' mainly focuses on a sufficient level of details, it is possible to say that the ninth property of Table I can be satisfied if its captured value provides diagnostic and sensitive information with a firm technical basis. With these definitions, it is possible to specify detailed explanations that provide practical use cases. In this regard, Table II summarizes practical examples of selective properties. It should be noted that, as the meanings of both 'Validity' and 'Intrusiveness' seem to be self-explainable, practical examples for these two properties were not explicitly given in Table II.

TABLE II. Practical examples to clarify the characteristics of each property

Property	Practical example
Diagnosticity	C1. Satisfied when the value is obtained through direct observation or expert survey. C2. Satisfied when a human performance measure provides specific standards.
Sensitivity	Time does not satisfy this property since it is difficult to define the exact range (0 to infinity). In contrast, a measure that specifies its range (e.g., 0 to 100) could provide the upper limit for visualizing the change rate of human performance with respect to the change of its value.

Reliability	Satisfied if its value has repeatability.
Objectivity	C1. Satisfied if a human performance measure is based on facts, phenomena, and behavior. C2. Not satisfied if an additional statistical analysis is required to understand the meaning of its value.
Simplicity	C1. Not satisfied when a human performance measure requires further knowledge of nuclear science, ergonomics, or psychology, or special data acquisition methods. C2. Not satisfied when a human performance measure requires special knowledge to understand the measurement results

Based on Tables I and II, it is possible to distinguish the characteristics of a human performance measurement. However, it seems that the definitions of the available dimensions are somewhat controversial because of a lack of detailed information clarifying the performance measurement criteria. Therefore, securing such detailed information would be helpful for practitioners who are responsible for the selection of human performance measurements. For this reason, in this paper, literature available from safety-critical industries such as the nuclear, aviation, and railway sectors were revisited in order to suggest the catalog of common properties with respect to the performance measurement criteria. In order to investigate the appropriateness of Tables I and II, Table III shows classification results for representative workload measurements: (1) Modified Cooper-Harper (MCH) scale, (2) NASA Task Load Index (TLX) and (3) Subjective Workload Assessment Technique (SWAT). From Table III, several insights can be obtained. For example, the characteristics of all measurements are very similar while 'Objectivity' seems to be the common limitation to be resolved. In other words, if the workload of human operators was measured by these three measurements, it is necessary to collect additional information to make up missing property (Objectivity). Accordingly, Tables I and II would be a good starting point to secure relevant human performance data to address diverse human factors issues. It is to be noted that, although many kinds of measurements are available, it is hard to specify a specific measurement that is preferable to a dedicated industrial sector. Indeed, this could be a goof rationale supporting the necessity of Table III that allows us to capture the pros and cons of individual measurement.

TABLE III. Property comparisons for representative workload measurements

Measurement	Diagnosticity		Sensitivity	Reliability	Objectivity		Simplicity	
	C1	C2			C1	C2	C1	C2
Modified Cooper-Harper scale [4]	Yes	N/A	Yes	Yes	No	No	Yes	Yes
NASA Task Load Index [5]	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Subjective Workload Assessment Technique [6]	Yes	N/A	Yes	Yes	No	No	Yes	Yes

ACKNOWLEDGMENTS

This work was supported by Nuclear Safety Research Program grant from the Korea Foundation of Nuclear Safety (KOFONS), funded by the Nuclear Safety and Security Commission of the Republic of Korea (No. RS-2021-KN052610).

REFERENCES

[1] L. Reinerman-Jones, S. Guznov, J. Tyson, A. D'Agostino, N. Hughes, *Workload, situation awareness, and teamwork*, NUREG/CR-7190, United States Nuclear Regulatory Commission, Washington D.C. (2015).

[2] J. M. O'Hara, W. F. Stubler, J. C. Higgins, W. S. Brown, *Integrated system validation: Methodology and review criteria*, NUREG/CR-6393, United States Nuclear Regulatory Commission, Washington, D.C.(1997).

[3] J. M. O'Hara, J. C. Higgins, S. Fleger, P. Pieringer, *Human factors engineering program review model*, NUREG - 0711, Rev. 3, United States Nuclear Regulatory Commission, Washington, D.C. (2012).

[4] M. L. Cummings, K. Myers, S. D. Scott, "Modified Cooper Harper evaluation tool for unmanned vehicle displays," *Proceedings of Conference on Unmanned Vehicle Systems Canada* (2006).

[5] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Sage CA: Los Angeles, CA, Vol. 50, p. 904-908, SAGE Publications (2006).

[6] G. B. Reid, T. E. Nygren, "The subjective workload assessment technique: A scaling procedure for measuring mental workload," *Advances in Psychology*, Vol. 52, p. 185-218 (1998).