Rating Scale for Evaluation Compliance with Learning Outcome Standards

Visitation teams will measure compliance with 7.**, 7.**, and 7.** separately using the rating scale below:

Rating	Description
Substantially Exceeds (The Program substantially exceeds the standard)	<u>All of the following are present</u> • Data from direct assessment measures that are of acceptable quality* clearly and consistently support the conclusion that the program is in compliance with the standard <center>And (additional criteria)</center>• Data from indirect measures that are of acceptable quality consistently support the conclusion that the program is in compliance with the standard • Contextual evidence (e.g., interviews, syllabi, assignments) consistently supports the conclusion that the program is in compliance with the standard • Both of the site visitors agree that triangulation of evidence clearly supports the conclusion that the program is in compliance with the standard
Exceeds (The Program exceeds the standard)	• Data from direct assessment measures that are of acceptable quality clearly and consistently support the conclusion that the program is in compliance with the standard <center>and</center>• Only two of the three additional criteria described in the rating of "5" are present
Met (The Program meets the standard)	• The preponderance of data from direct assessment measures that are of acceptable quality support the conclusion that the program is in compliance with the standard <center>and</center>• Only one of the three additional criteria described in the rating of "5" is present
Not Met	The preponderance of evidence indicates that the program is not in compliance with this standard, or the evidence about compliance is not sufficient to warrant the conclusion that the program is in compliance with the standard. Data from direct measures do not provide compelling evidence

(The Program does not meet the standard)	of compliance. Some indirect measures may indicate compliance.
Serious noncompliance (The program is seriously out of compliance with the standard	Sources of evidence clearly and consistently indicate that the program is not in compliance with this standard

*"Acceptable quality" means that content-related evidence of validity is present. More formally, a measurement tool ("metric") is of acceptable quality if inferences made about students' achievement of learning outcomes based on scores from the measurement tools are justifiable. Site visitors should consider content-relevance and content-representativeness in making judgments about content-related evidence of validity. As such, the metrics should a) include quality indicators of the relevant dimensions of the learning outcome being measured (content relevance), and b) be representative of the major facets of the learning outcome that is the target of the measurement.

 A simple example is a test of comprehensive elementary arithmetic skill that might be administered to grade-school children. Such a test should present arithmetic problems from all four dimensions of arithmetic: addition, subtraction, multiplication, and division. If one of these dimensions was not assessed (e.g., no division problems were included on the test), the test would lack content representativeness. In this case, it would not be justifiable to make an inference about students' comprehensive arithmetic skills if the test lacked items from one or more of the four dimensions of arithmetic. That test did not generate evidence of skill in one of the essential four dimensions. Similarly, if the items on the test were statements about students' attitudes about mathematics instead of being arithmetic problems to be solved, the test would lack content relevance. That test might (or might not) be an acceptable measure of **attitude** toward arithmetic, but it would not be justifiable to make an inference about a given student's arithmetic **skill** based on responses to these attitudinal questions.

In the context of COA assessment, indicators should, in the site visitor's judgment, be content-relevant and content-representative. A measure of operations management skill (whether it be a written exam, direct measure of performance, or other measure) would lack content-representativeness if that test was limited to questions about only one dimension; financial

management, for example. Other dimensions of operations management must also be assessed (e.g., planning, organizing, leading, directing, controlling). The indicators must also be well-written and clearly presented, and procedures for converting responses or behaviors directly observed to scores must be unambiguous. On a written exam, items should be written clearly and succinctly, ambiguity should be minimal, student's task should be clear (e.g., choose one of four options on a multiple choice test), and a constant rule for converting responses to total scores should be used. The total score might, for example, be the sum of correct responses, or perhaps some items might be weighted differently (a greater number of points awarded for a correct response). If the measure is of directly observed behavioral performance, the behaviors observed should be expected to be recorded similarly by different observers (inter-rater agreement), and the range of behaviors that are recorded for scoring must represent the range of dimensions of the concept. It would not be acceptable, for example, to evaluate only a budget assignment and then make an inference about students' operations management learning outcomes based on the relative quality of different budget projects.