EVALUATION THEORY AND PRACTICE AS APPLIED TO SECURITY EDUCATION
An Overview

Melissa J. Dark
Purdue University

Abstract: This paper will overview general evaluation purposes, elements, and steps for designing an evaluation in order to provide foundational information that can be used to conduct an evaluation of any security awareness, training, or education programs. An example of evaluation principles discussed in this paper as applied to an information security education program has been provided in the appendix. This paper is a tool for individuals who have little to no formal training in educational evaluation.

Key words: evaluation, evaluation design, formative evaluation, summative evaluation, measurement, metrics, program logic, validity, reliability.

1.   INTRODUCTION

The information assurance environment is shaped by new technologies, unknown threats, increasing vulnerabilities, a national security workforce crisis, and a lack of sufficient security education. The lack of human resource is the most critical because it underlies each of the other issues. In response, new security education initiatives and programs are rapidly being established to expand our human resources capacity. While several initiatives intend to address this need, there will be many unasked and unanswered questions about the impact of such programs and the extent to which these initiatives succeed. A common misconception in education is that evaluation does not need to be seriously considered until the end of the awareness, training, and/or education program. However, this fallacy contributes to educational programs that are less effective than needed and less efficient than they had to be. This paper will overview general

evaluation purposes, elements, and steps for designing an evaluation in order to provide foundational information that can be used to conduct an evaluation of any security awareness, training, or education programs. An extensive example is provided in the appendix wherein the principles presented in this paper are applied to a security education program.

2.   DEFINING EVALUATION

What is evaluation? Evaluation is concerned with the investigation of the worth or merit of an object and conducted for purposes of improving, informing, and/or proving (1). In the broadest sense, evaluation "includes all efforts to place value on events, things, processes, or people" (3). Evaluation is both the process of conducting an evaluation or a program as well as the resulting product, e.g., an evaluation report. Evaluation of educational programs utilizes social research procedures to systematically investigate the effectiveness of the programs. Evaluation spans the life cycle of any program including the diagnosis of the social problems being addressed, conceptualization and design, implementation and administration, outcomes, and efficiency (3).

Evaluation is generally conducted for the purposes of informing, improving, and/or proving. Evaluation that is conducted to identify needs and to understand the nature of the need(s) is known as needs evaluation/analysis. Needs analysis is usually done prior to program development and includes investigation that both quantifies and qualifies the nature of the need. Evaluation that is undertaken to furnish information that will guide improvement to the educational program or activity is known as formative evaluation. For example, a formative evaluation might be intended to a) help management improve community acceptance of a security awareness program, b) improve a security training program through the use of a new laboratory activities, or c) help an organization improve the cost effectiveness of a security education program by converting it to self-paced, computer-based instruction. Formative evaluation is conducted in the early stages of developing an educational product or program or in the early stages of implementation when there is still sufficient time to make use of the evaluation results. In contrast to formative evaluation, evaluation undertaken to render a final judgment is known as summative evaluation. Examples of summative evaluation might include a) judging the effectiveness of a faculty development program that intends to develop faculty knowledge and skills; or b) show the extent to which security vulnerabilities were reduced as a result of a new training program. Summative evaluation is conducted at the end of the educational program, activity, module, etc.

3.   THE EVALUATION SYSTEM

The evaluation of a program will only produce relevant, useful, and accurate information if the evaluation 'system' is designed to ensure that this happens. We'll start by looking at elements common to all evaluation systems and proceed to an overview of steps that can be followed to design any evaluation.

Evaluation systems are comprised of metrics and measurement systems. These terms are often used interchangeably or misused in education, so it is worth clarifying the differences and relationships. The purpose of the measurement process is to establish the degree or amount to which a characteristic, trait, or feature exists. The measurement process almost always includes a measurement instrument of some sort that produces a measure (e.g., a score, rating, ranking, etc.). The result of the measurement process is a measure that serves as an indicator of the desired trait, skill, ability, etc. Most measures are meaningless until they are contextualized by a measurement system (e.g., a scale of some sort). A measure that is contextualized by a measurement system is a metric. The measurement system is critical in evaluation because it allows program mangers, policy makers, evaluators, and other stakeholders to ascertain the degree or extent to which the desired property, trait, process, etc., exists across programs, across time, relative to expectations, etc. Measurement always involves three common steps: 1) identifying and defining the quality or attribute that is to be measured, 2) determining a set of operations by which the attribute may be made manifest and perceivable, and 3) establishing a set of procedures or definitions for translating observations into quantitative statements of degree or amount (4). Error can be introduced into the measurement process at each of these steps; erroneous measures and measurement lead to inappropriate evaluations.

While measurement and metrics are an essential part of evaluation, evaluation goes beyond measurement. Evaluation is the systematic process of collecting, analyzing, and interpreting information (2) in order to render judgments. As previously mentioned, these judgments might be in regard to improvement (formative evaluation) or to render final judgment (summative evaluation). In addition to measurement error, error can also be introduced through a poorly designed evaluation. The remainder of this paper outlines recommended steps and associated considerations for designing an effective evaluation. An example of each of the following steps as applied to an information security education program has been provided in the appendix.

4.   STEPS FOR DESIGNING AN EVALUATION

Following is a discussion of five steps to consider when designing an evaluation. I will discuss each step in detail and have provided a detailed example of each step as applied to security education in the appendix.

The first step is to determine whether the purpose of the evaluation is needs analysis, formative, or summative. Needs analysis, formative, and summative evaluation questions are distinctly different in nature as exemplified by the evaluation questions shown in table 1.

Table 1: Needs Analysis/Formative/Summative Evaluation

Needs Analysis	Formative	Summative
What learning goals are not being met?	What problems are students having learning the content and why?	How well were goals/objectives met?
How/why are current programs insufficient?	Are the awareness campaigns having the desired impact and why/why not?	Did any changes made have the desired impact?
Are current programs adaptable for a new/different learner population?		Is it worth teaching again?
		Was it cost effective?
What changes are needed to programs to meet the needs of a new learner population?	How should the curriculum be updated to meet industry needs?	To what degree did the training program alter employee behavior and mitigate our security risk?
	Where is this program least cost effective and what can be done to improve cost effectiveness?	How many students passed the exam?

Lack of proper focus will inevitably lead to an inappropriate evaluation; when this happens decisions based on the evaluation are bound to be error-ridden. In addition to poor focus, another grievous mistake is to attempt to take one type of data (e.g., data that were collected for a formative evaluation) and use these data for another purpose (e.g., to answer a summative evaluation question). Data for the most part are meaningless. Data only become meaningful when used in a context and for a purpose; the appropriate reuse of data is dependent upon the degree of alignment between the original and the new context and purpose. The evaluator needs to clearly determine the purpose of the evaluation and adhere to it throughout the evaluation process.

The second step is to frame the evaluation. This step is critical to beginning the evaluation properly. In a way, conducting an evaluation is like building a house; one should not build a house without a design that

specifies what goes where and why. Likewise, when conducting an evaluation, a design is needed that serves as a blueprint for what the evaluation will accomplish. The design of the evaluation should reflect the purpose (needs analysis/formative/summative) of the evaluation. When the evaluation is a needs analysis, the object of the evaluation is to ascertain the nature and scope of the need so that programs can be developed that will address the need, i.e., fill the gap. By nature, a needs analysis is an evaluation of a state or condition. The purpose of a formative or summative evaluation is to evaluate a program that has been established in response to a known need. By nature, formative and summative analyses are evaluations of a program, initiative, etc., developed in response to the needs. Therefore, formative and summative evaluation should also reflect the framework of the security awareness, training, or education program being evaluated.

The framework of any program can be determined by identifying the program logic, which consists of antecedents, transactions, and outcomes, as well as the underlying beliefs, assumptions, and theories (figure 1).
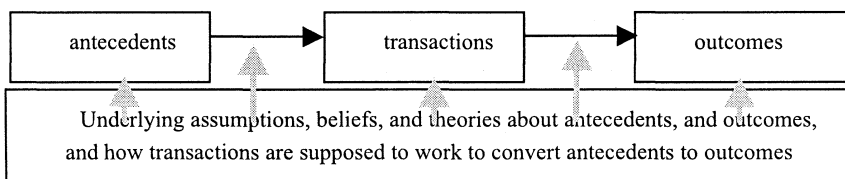

Figure 1: Program Logic Model

Antecedents are the inputs into the program. Outcomes are the goals and objectives that the program is trying to achieve. Transactions are the activities within the program that convert antecedents into outcomes. Educational antecedents, transactions, and outcomes can include policies, programs, curricula, courses, instructional software, instructional materials, delivery systems, learners, and so on. By identifying underlying assumptions, beliefs, and theories about how transactions are supposed to work to convert antecedents into outcomes, the evaluator can fully define the conceptual framework of the program. There is not a standard process for identifying and documenting program logic, however, there are several models that can be utilized to guide this process; some of which are the Provus Discrepancy Model, the CIPP (context, input, process, product) model, Stake's countenance approach, Eisner's connoisseurship and criticism approach, Stake's illuminative evaluation, and so on. Which model is most appropriate depends upon several factors including, but not limited to the purpose of the evaluation, the evlauand (the program, persons, etc., being evaluated), the audience for the evaluation, the required degree of

rigor, and the existence of predetermined and widely accepted standards. Once program logic is manifest, the evaluator is in a position to 1) use this information as a blueprint for determining the extent of program implementation, 2) identify gaps and contradictions in the conceptual framework and causal links, 3) identify unrealistic expectations for the program, and 4) reveal possible side effects of the program.

After critical aspects of the program have been defined and documented, the next step is to determine evaluation questions, i.e., what questions is the evaluation seeking to answer. When drafting evaluation questions, you will want to attend to the tone and style of the questions, as well as the content focus of the questions. The tone and style of the evaluation questions reflects whether you are conducting a needs analysis, formative, or summative evaluation. The content focus of the questions reflects program logic (antecedents, transactions, and outcomes, and underlying assumptions, beliefs, and theories).

I like to use two levels of evaluation questions. The first level question is a comprehensive evaluation question that cannot be answered in and of itself, such as the example provided in figure 2. Level one questions guide the evaluation, are generally timeless, and aligned to level two questions.
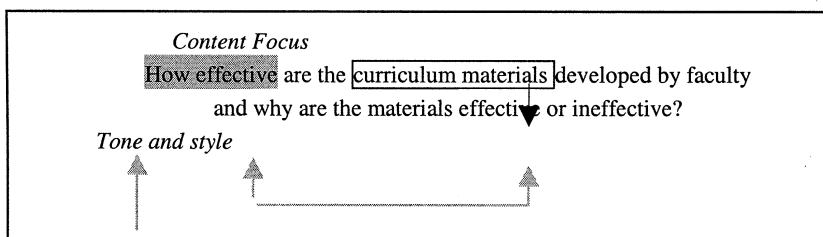
Figure 2: Level One Evaluation Question

Second level (level two) evaluation questions are more specific. The collective answers to the second level questions paint a picture that answers the first level question. Level two questions are based on underlying assumptions, beliefs, and theories. Level two questions often change as the program is modified and improved, but remain aligned to level one. Good level two questions: 1) indicate specific criteria of merit, 2) orient you toward a measure(s) of performance outcomes, 3) focus on only one aspect of a goal, 4) are designed with thought toward the information that could be collected to inform them, 5) consider any subpopulations that need to be examined separately, 6) are actionable so that they guide improvement or decision-making, 7) are reliable (i.e., can be measured by a variety of people and across time, and 8) are comparable (i.e., measures developed from level

two questions should be able to be compared to measures developed from similar questions in the past).

Following are sample statements about effective curriculum materials that demonstrate guiding assumptions, beliefs, and theories: 1) curriculum materials should be current in order to provide students recency of content, 2) curriculum materials should have content validity. 3) curriculum materials should have construct validity, 4) curriculum materials should include opportunities for active learning, 5) curriculum materials should specify intended learning outcomes. These statements can easily be transformed into level two evaluation questions as shown in figure 3.

To what extent are the curriculum materials current?
To what extent do the materials have content validity?
To what extent do the materials have construct validity?
To what extent do the materials provide opportunities for active learning?
To what extent do the materials specify intended learning outcomes?

Figure 3: Level Two Evaluation Questions

With evaluation goals identified and questions formulated, the fourth step is to figure out what information is needed to answer evaluation questions. More specifically, measures and a measurement system need to be defined. The evaluator should identify and define the qualities or attributes that are to be measured, determine indicators of the attributes that are manifest and perceivable, and 3) create, adopt, and/or revise instruments to collect data. While a comprehensive discussion of measurement is not possible in this paper, there are a few key points I want to reiterate. The qualities or attributes that are to be measured should be derived from the program logic model. By now, these qualities and attributes should be evident in your level two evaluation questions. If these qualities and attributes are not evident, then you should go back to step three. If you do not know what you qualities you are assessing, then it will be impossible to know what measure will serve as indicators of these qualities. Good indicators are 1) oriented toward a measure of performance, 2) specific, 3) operational definitions of the attribute you are trying to assess, 4) theory-driven, 5) valid, and 6) reflective of the program goals, questions, elements, assumptions, beliefs, etc. Indicators can be viewed incrementally or as a fixed target. For example, in some cases it makes sense to report an increase in pass rate on the national examination (pass rates increased by 1% every year for 4 years), but in other situation it makes more sense to report a fixed target (all students passed at or above proficiency levels).

When measuring improvement incrementally, it is necessary to establish a baseline as a starting point. In this instance, you will need some type of comparative design in order to detect change, such as a cyclic or pre/post design. While the observed change is often sufficient evidence of program effects, it is possible that observed effects are due to other factors, such as maturation, other conditions, another program happening at the same time, and so on. If evidence based on a pretest/posttest design is not conclusive enough, then it might be necessary to use a pretest/posttest comparison group design. The pretest/posttest comparison group design collects information for both groups (treatment and control) both before and after the program. With this design, both the treatment and the control group will experience change due to maturation, testing, history, etc., making it possible to attribute differences between the two groups to the security education program. The addition of a control group will require that the available participant group be larger. As the need for rigor increases, the evaluation design becomes more sophisticated and the resources required to conduct the evaluation also increase. Occasionally, the measurement step is complete once sufficient indicators have been identified. For example, if student test scores on a national certification examination are an indicator of well-prepared graduates and if one indicator is sufficient, then it might not be necessary to create, adopt, or revise an instrument to collect indicator data. However, in many evaluation situations, the required indicators are a combination of existing measures as well as new measures making it necessary to create, adopt, and/or revise an instrument(s).

The fifth step in the evaluation design process is to establish a systematic method(s) for collecting the information. There are many different methods for collecting information. Again, a comprehensive discussion is not possible; instead I will provide tips and techniques for improving data collection and overview various data collection methods.

Pilot testing is a trial run of the instruments that you plan to use and can be invaluable in collecting 'good' data. Pilot testing gives the investigator an opportunity to identify and address instrument error, can save you time and money, and give you an idea how long data collection might take. Pilot testing can range from sending the instrument to a few of the intended users and then calling them to discuss the questions to actually conducting a pilot study. The effort put into pilot testing usually is dependent upon availability of resources and required degree of rigor in the evaluation. When pilot testing, your goal is to discern sources of measurement error in the instrument. Measurement error is an attribute of the instrument that keeps the user from correctly and completely responding, which in turn biases your data.

Triangulation is the utilization of many indicators and sources of data. Many sources are often better than one when looking for evidence. For

example, if you are looking for evidence of well-prepared students and your students obtain above average national certification examination scores, reports from alumni that they were well-prepared for their first job in the occupation, and data from employers that graduates are well-prepared, you have in essence controlled for measurement error and increased the degree of confidence with which you can ascertain that your students are in fact well-prepared. The more sources of evidence you have, the more faith you can have in a result or finding.

Sampling is sometimes required in an evaluation study. This is usually a result of a large population and/or limited resources. Depending upon the nature and purpose of the evaluation, the evaluator will need to make decisions about sampling. When attempting to make generalizations about the entire population based on the sample, principles of sampling, including representativeness and sample size, will need to be considered. When investigating unique cases or seeking detailed information about transactions or outcomes, purposive sampling techniques will be more appropriate.

There are a variety of different methods for collecting data, e.g., surveys, tests, etc. Regardless of the method used to collect data, evaluators need to attend to the validity and the reliability of their instruments. Validity refers to the degree to which an instrument measures what it is supposed to measure. Due to the fact that instruments are usually designed with a specific purpose, the validity of the instrument is a function of that purpose. Validity is an important issue both when you create your own instrument, when you adopt another instrument for use, and when you make interpretations based on results. Validity is not a number, it is a matter of degree. A test is not valid or invalid; it is more a matter of having high validity, moderate validity, or low validity. Reliability refers to the consistency of the evaluation results. If we give a test to a group on two different occasions and receive the identical or very similar results, then the instrument is fairly reliable. If two evaluators conduct a focus group and report very similar findings, then the there is said to be rater or interviewer reliability. While reliability is conceptually distinct from validity, there is a relationship between the two. Reliability is a necessary, but not sufficient condition, for validity. You can have an instrument that measures the wrong thing (invalid), but does so with great consistency (reliable). It is possible to have reliability without validity, but it is not possible to have validity without reliability. Both validity and reliability are needed to base interpretations on the data that is collected.

When designing instruments to collect information that will be used in an evaluation, you will also want to consider the characteristics, advantages, and disadvantages of various methods for collecting data. Surveys can vary along a number of dimensions such as how questions are asked and how the

instrument is delivered. Questions can be asked in a closed-ended versus open-ended manner. Surveys can be mailed, distributed in person, conducted over the telephone, or in a group setting. Determining what to ask on a survey and how to ask it is not an easy task. There are entire books devoted to this topic. For the sake of this paper, here are a few basic tenets. First, identify exactly want type of information you want responders to provide. Generally, questions can be classified into two types: 1) behavior or attribute questions, and 2) beliefs or attitudes questions. For example, in a study designed to assess instructor effectiveness, you need to decide in advance what type of information you want/need to collect. If you want to know about behavior, then the following questions would be appropriate: 1) My instructor utilizes a variety of strategies to teach this class, or 2) My instructor allows me to ask questions and gives me timely feedback. However, if you want to know beliefs, such as how students feel about the course, then the following questions might be appropriate: 1) In your opinion, how effective were the instructional strategies at helping you learn, or 2) I believe that this instructor is among the best I have ever known. Second, think about exactly what you want to know and ask it as directly as you can. Be specific, use simple words, avoid jargon, and keep it as short as possible. Third, eliminate bias from your questions. Bias in questions leads your respondent to answer in the positive light, regardless of whether it is accurate or not. Take for example the following questions: 1) Do you agree that instructor's who offer test review sessions are more effective than instructors who do not, and 2) Graduates of the National University program are generally more prepared today than they were 3 years ago. Have the graduates that you have hired been more prepared?

As opposed to surveys, tests are instruments designed to assess learning. Instruments can vary from written tests, oral examinations, to performance evaluations. Within those categories, tests often include a variety of formats for items, including multiple choice, true-false, short answer, and essay. Determining what to include on a test is not a simple procedure. There are some basic tenets that can be used to construct tests that are more valid and reliable. First, identify the purpose of testing. Tests can be used for a variety of reasons including placement, formative evaluation, diagnostic testing, summative testing, or prediction. Placement testing will focus on a minimum amount of knowledge, whereas formative evaluation will focus on a predetermined segment of instruction, and summative testing will focus on an entire course or program. Second, develop a set of test specifications. A set of specifications will ensure that the test has content validity. To build a set of specifications, you will need to delineate intended learning outcomes, corresponding course content, and prepare a specification chart that details and weights outcomes by content area. Third, select appropriate items. Types

of items can be classified as two general types: objective items and essay items. Objective test items include any type of question where response options are supplied, and essay items include any type of question where the learner must generate and present the answer in some format. Objective items are efficient for measuring facts, understanding, and on a limited based more complex outcomes, whereas essay items are appropriate for more complex outcomes. Scoring is relatively easy on objective test items and high reliability is more likely. Scoring on essay items tends to be slow and subject to more inconsistencies, thereby becoming a potential threat to reliability.

Interviews are well suited to collecting detailed information from a smaller group of participants. Interviews are generally more time-consuming and costly than a survey, but well written interviews can produce a wealth of information. Interviews can range from very structured, where the questions are written out and are to be covered verbatim, to extremely unstructured (where a list of general topics to be covered is followed). The decision on which type of approach to use depends upon the nature of the evaluation and the type of information that you are trying to collect.

Focus groups are an effective way to collect large amounts of detailed information on a restricted budget. A focus group is like a group interview and can also be conducted using a structured or unstructured approach. However, in focus groups it is difficult to ensure that you get the input of all participants. When conducting a focus group, you should: 1) identify what information you are trying to collect, 2) create a focus group guide to follow, 3) be prepared to take copious notes….you might even want an audio tape so you can capture it all, and 4) seek out the opinion of quiet participants.

Sometimes the best way to collect information about people's behavior is to watch them. Observation allows you to collect information without burdening the person providing the information. However, observation is limited in its use to what you see. Observation is not an effective method for collecting information about attitudes, feelings, etc. When conducting an observation, you should: 1) identify what information you are trying to collect, 2) determine specifically what behaviors will serve as indicators, 3) create a checklist for use during observation, and 4) train observers on what to look for and how to accurately record observations on the checklist.

Evaluating any educational program can be a challenging endeavor and the same will be true in security education. As more and more institutions, organizations, schools, and programs launch security education programs in an attempt to meet needs that are emerging in a rapidly changing environment, evaluation will be more difficult, yet even more important in order to ensure that programs are having the desired impact. The need to embrace this challenge will be even greater. The questions facing us are many and varied. How do we assess educational programs that vary in: 1)

types of students served, 2) curriculum areas, 3) facilities, 4) policies, 5) stakeholders served, and so on? My hope is that this paper can guide users from a variety of institutions and programs through the process of educational program evaluation with the goal of raising the quality of security education evaluations with the subsequent goal of bettering security education programs.

5.   APPENDIX

5.1   Step 1: Determine the Purpose of the Evaluation

5.1.1   A Capacity Building Program in Information Security Education

Micheal Hanes is the Director of the Information Security Consortium. The goal of the Information Security Education Consortium was to develop, delivery, disseminate and evaluate a multidisciplinary program to develop faculty expertise and capacity in teaching Information Assurance and Security at the undergraduate and graduate levels. By doing so, they hoped to improve the capacity and capability of higher education to address the critical shortage of Information Assurance and Security specialists.

In order to develop faculty expertise and capacity in teaching Information Assurance and Security, they had the following objectives:

* Develop a Core of Multi-Disciplinary Faculty Mentors

* Experts will be recruited to serve as mentors from academia, industry, and

* Mentors will attend a mentor workshop to structure guidelines for the design and development of curriculum materials and the faculty development institute.

* Faculty mentors will work with recruited faculty for two years to facilitate integration of information assurance and security into their academic programs.

* Develop Curricular Materials to be used in Faculty Development Institute

  * Materials will be current in information assurance and security subject matter.

  * Materials will be current in educational methodology for effective undergraduate and graduate education.

  * Educational methodology for using the materials in undergraduate and graduate programs will be explicitly addressed so that participating faculty understand and know how to use the materials most effectively.

- Develop Subject-Matter Expertise of Participating Faculty through a Faculty Development Institute led by Faculty Mentors

  - Institute will utilize current educational methodology for effective faculty development.

- Dissemination of Curriculum Materials

- Curriculum materials will be made available through publishers, presentations at professional conferences such as NCISSE, and NISSC, and through related journals.

In addition to leading the project, Michael was responsible for assessing the effectiveness of the program. One of the first things that Michael asked himself was why was assessment necessary in the first place? Michael realized that different people had different reasons for wanting an assessment. The consortium had certain reasons, the mentors had other reasons, and the sponsor had yet other unique interests and reasons.

Upon talking with these groups, Michael learned that the consortium wanted to find out to what extent their program was building capacity, how many educators were participating, which activities were most and least effective, and how these activities were contributing to capacity building. The mentors specifically wanted to find out to what degree their materials were utilized in other classrooms and how, their effectiveness, and suggestions for improvement. The sponsor was interested in knowing the cost benefit of the program as well as ways to improve the cost benefit ratio.

So why was Michael assessing this program? The answer to this question had two major parts:

- To find out the extent to which the program is producing the intended results, and

- To find out where is in performing as planned, where it is not, and methods for improvement so that it does perform as planned in the future.

5.2 Step 2: Frame the Evaluation

5.2.1 The Conceptual Framework for Information Security Education Program

5.2.2 Antecedents

- Expertise of those chosen to be mentors in information and security assurance content areas

- Faculty attendees will want and need to integrate topics into their curriculum

- Curriculum development model

5.2.3   Transactions

- Mentor workshop to teach mentors curriculum development format
- Development of curriculum materials to integrate information assurance and security into other curriculum areas
- Faculty members attend institute and work within their own fields as well as with experts in the field of information assurance and security
- Mentors assist faculty in integration.
- Materials are integrated into existing curricula
- Materials are disseminated to a wider audience

5.2.4   Outcomes

- Mentors enhance their curricula with materials developed during institutes
- Effective curriculum materials
- Curriculum materials disseminated through various channels to a wide audience
- Faculty members integrate information assurance and security information into their curricula
- Students in core courses gain experience and knowledge in the field of information assurance and security
- Students are able to use skills in the field of information assurance and security in order to enhance their work and increase their potential value in the job market

5.2.5   Underlying Assumptions, Beliefs, and Theories

Current instruction in information assurance and security is insufficient to support the workplace need for skilled workers in this field. Institutions and faculty see a similar critical need for this type of instruction as the proposal's authors. There is an established link between instruction in information assurance and security and increased use. Participating institutions will pay for half of the costs incurred by faculty attending the development institute. A faculty institute is the best method for serving the proposed clientele. Instructors will be able to or willing to modify their curricula to reflect information security instruction. Institutions see a need

for a change in cross-curricular preparation in this area of technology. Development of curricular materials will have an impact at the classroom level. The field of information security has far-reaching implications in all fields. Faculty mentoring is an effective way to facilitate curriculum changes. Faculty and curriculum development will increase capacity.

5.3 Step 3: Determine Evaluation Questions

5.3.1 Assessment Questions for the Information Security Program

Michael had already asked himself why different stakeholders wanted an assessment and what they wanted from the assessment. He had come with two major reasons for the assessment. He had also reviewed the conceptual framework of the program and identified the intended antecedents, transactions, and outcomes. With the broad purpose in mind and the key components of the program laid out, Michael was able to frame key assessment questions that were well grounded. The following is a result of his work:

Program Goal	Level 1 Questions	Level Two Questions
Curriculum development	How effective is the curriculum development process and the resulting materials?	To what extent are mentors experts in the field? To what extent did the mentor workshop specify curriculum development criteria. To what extent were the criteria based on instructional theory? To what extent do the curriculum materials specify learning outcomes? To what extent are the curriculum materials valid and usable?
Faculty development	How effective is the institute at helping educators integrate IAS into their curriculum?	To what extent do faculty attendees gain knowledge and skills in IAS topics. To what extent do faculty attendees gain knowledge and skills in using curricular materials? To what extent do faculty mentors facilitate successful integration with faculty attendees?
Curriculum dissemination	How effective are dissemination efforts? Why or why not?	To what extent do faculty attendees integrate materials into their curriculum? To what extent are materials made available and used outside the institute?
Cost benefit	How effective is the program at maximizing benefit of cost investment?	What are the most/least beneficial aspects of the program and why? What is the cost investment? How can the program cost/benefit be improved?

Michael reviewed the questions to determine if there was anything he was leaving out based on his priorities and those of other stakeholders. He determined that the goals and questions he had were valid questions and went on to determine the design of the assessment.

5.4   Step 4: Determine What Information is Needed to Answer Evaluation Questions

What Information Did the Information Security Education Consortium Need to Collect?

Level Two Questions	Indicators	Existing	TBD
To what extent are mentors experts in the field?	Number and type of publications of mentors	X	
To what extent did the mentor workshop specify curriculum development criteria?	Number and quality of curriculum guidelines		X
	Usability of guidelines		X
	Adherence to guidelines		X
To what extent were the criteria based on instructional theory?	Congruence between learning outcomes and appropriate instructional strategies		X
To what extent do the curriculum materials specify learning outcomes?	% of materials that bear evidence of learning outcomes		X
To what extent are the curriculum materials valid and usable?	Faculty ratings of validity and usability		X
	Faculty use		X
To what extent do faculty attendees gain knowledge and skills in IAS topics?	Test scores		X
	Integration rates		X
To what extent do faculty attendees gain knowledge/skills in using curricular materials?	Test scores		X
	Integration rates		X
To what extent do faculty mentors facilitate successful integration with faculty attendees?	Amount of time mentors work with faculty		X
	% of integration		X
To what extent do faculty attendees integrate materials into their curriculum?	% of integration		X
	Length of integration		X
	# of students exposed		X
To what extent are materials made available and used outside the institute?	# and methods of dissemination		X
	# of requests for materials		X
	reported use of materials		X
What are the most/least beneficial aspects of the program and why?	Outcomes of materials		X
	Outcomes of faculty development		X
	institute and Outcomes of curriculum dissemination		X
What is the cost investment?	Cost per activity		X
How can cost/benefit be improved?	Cost/benefit ratio		X

5.5   Step 5: Method(s) for Collecting Information

How Did the Information Security Education Program Collect Information?

Michael listed possible ways to collect information pertinent to his measures and came up with the following:

To what extent are mentors experts in the field?	Number and type of publications of mentors	Records
To what extent did the mentor workshop specify curriculum development criteria?	Number and quality of curriculum guidelines Usability of guidelines Adherence to guidelines	Records and Materials Survey Records and Materials
To what extent were the criteria based on instructional theory?	Congruence between learning outcomes and appropriate instructional strategies	Records and Materials
To what extent do the curriculum materials specify learning outcomes?	% of materials that bear evidence of learning outcomes	Records and Materials
To what extent are the curriculum materials valid and usable?	Faculty ratings of validity and usability Faculty use	Survey Survey/observation
To what extent do faculty attendees gain knowledge and skills in IAS topics?	Test scores Integration rates	Test Survey/observation
To what extent do faculty attendees gain knowledge/skills in using curricular materials?	Test scores Integration rates	Test Survey/observation
To what extent do faculty mentors facilitate successful integration with faculty attendees?	Amount of time mentors work with faculty % of integration	Records Survey/observation
To what extent do faculty attendees integrate materials into their curriculum?	% of integration Length of integration # of students exposed	Survey/observation Survey Survey
To what extent are materials made available and used outside the institute?	# and methods of dissemination # of requests for materials reported use of materials	Records Records
What are the most/least beneficial aspects of the program and why?	Outcomes of materials Outcomes of faculty development institute and Outcomes of curriculum dissemination	Observation/Survey/ Records Observation/Survey/Records Observation/Survey/Records
What is the cost investment?	Cost per activity	Records
How can the program cost/benefit be improved?	Cost/benefit ratio	Records

REFERENCES

1. Evaluation Center. (May 10, 2000). <u>About evaluation</u>. (On-line). Available: http://www.wmich.edu
2. Gronlund, N. (1985). <u>Measurement and evaluation in teaching</u> (5th ed.). New York: Macmillan.
3. Rossi, P., Freeman, H., Lipsay, M. (1999). *Evaluation: A Systematic Approach*. 6th edition. London: Sage Publications.
4. Thorndike, R., & Hagan. E. (1977). <u>Measurement and evaluation in psychology and education</u> (4th ed.). New York: Wiley.