TEST DEVELOPMENT

UNIT EXERCISE 3

3/4/2010

R561: Evaluation in Instructional Development Process

[redacted]

TABLE OF CONTENTS

EXECUTIVE REPORT PREPARED FOR BLOOMINGTON UNIVERSITY BY HELIOS

HELIOS' UNDERSTANDING OF CLIENT'S (BLOOMINGTON UNIVERSITY) NEEDS

Helios, a consulting firm is recruited by the Instructional Systems Technology (IST) department at Bloomington University in Bloomington, Indiana. Helios is recruited to create test questions on R561, an Evaluation course in the Instructional Development Process and make recommendations for improvements.

R561 is a newly introduced course and has been deemed mandatory for all Master's level students graduating the program with IST as their major. Dr. James Jones is the instructor for this course and the Project Sponsor. Dr. Jones believes that Kirkpatrick's Four Levels of Evaluation is fundamental to R561 and since the course has been newly introduced it was the department's decision to assess students learning early on and refine the course based on recommendations from the consultant. Furthermore, since most organizations and practitioners have adopted "The Four Levels" Dr. Jones firmly believes that the graduating class of 2010 needs to be thoroughly conversant with Kirkpatrick's Evaluation Model.

COURSE R561 OUTLINE

- ❒ The course is distributed over a 15-week schedule
- ❒ Kirkpatrick's Model of Evaluation will be completed on Week 13
- ❒ Final class project due on Week 15

COURSE R561 GOAL

- ❒ Identify the basic concepts and terminologies associated with evaluation
- ❒ Gain mastery of Kirkpatrick's Model of Evaluation
- ❒ Understand the role of evaluation within the ISD and PT process

ASSUMPTIONS

1. Since the test questions are being developed by Helios when the students have completed only three weeks of the course, it was agreed that the test would be constructed to test the first two levels of (revised) Bloom's taxonomy i.e. Remembering (Level I) and Understanding (Level II) (Forehand, 2005).
2. It is assumed that some students will reach Level 4 (Analyze) (Forehand, 2005), the ability to determine, differentiate, organize and attribute how the parts relate to one another and to an overall structure of Kirkpatrick's Model of Evaluation (on Week 13) and some students may attain Level 5 i.e. the ability to critique. However, Level 6, the ability to create may not be attained at all by many or most of the students.
3. Given the following constraints of the R561 course such as time, varying levels of aptitude and that the course has been taught for only three weeks, it is further assumed that:
    a. **Knowledge** will be learned minimally (Kirkpatrick & Kirkpatrick, 2006). It is further assumed that since students have been exposed to only 3 weeks of the Kirkpatrick Model, most

students will be relegated to Level 1 of Bloom's Taxonomy and some may have moved to Level 2 due to varying levels of student aptitudes.

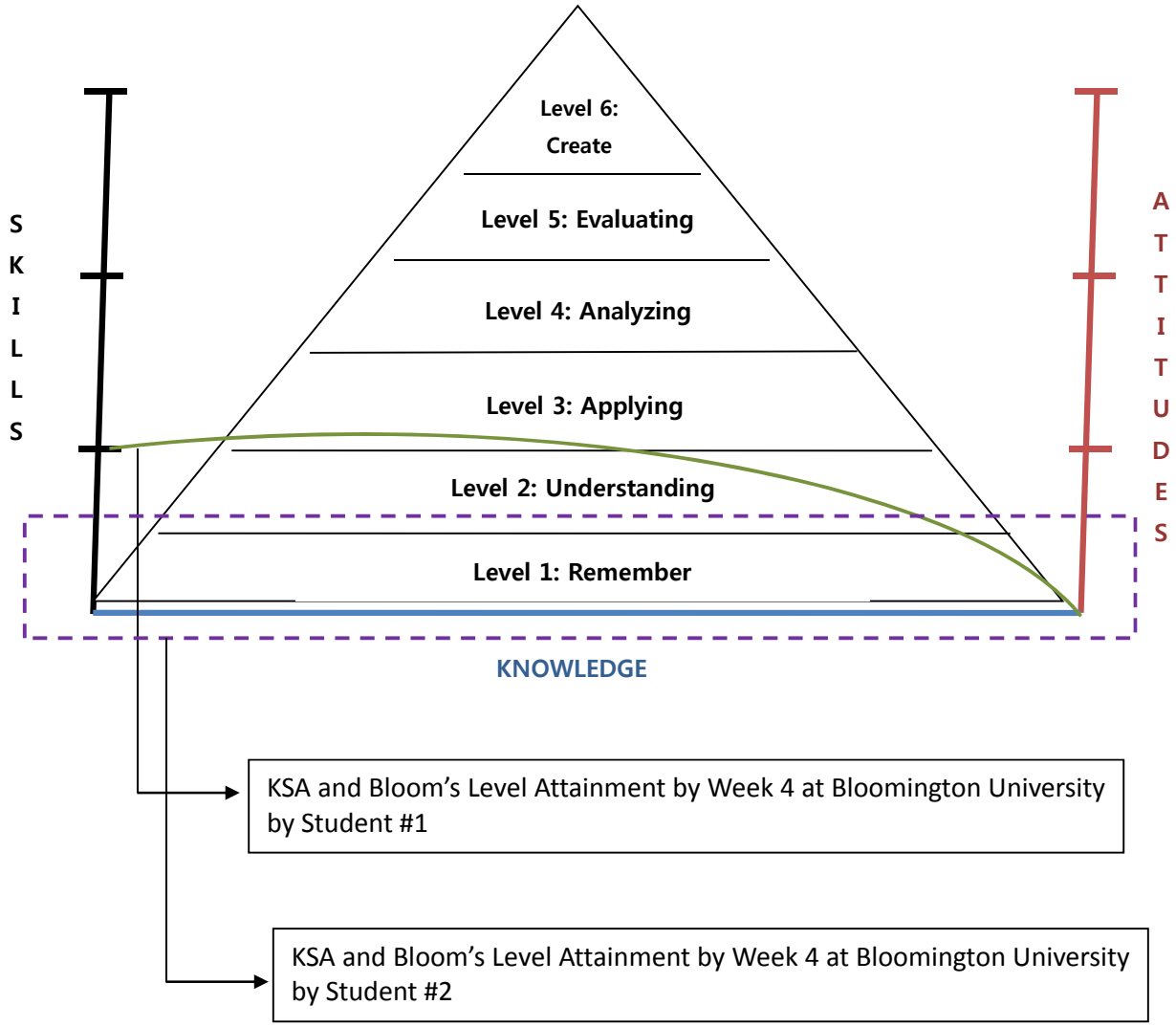
b. **Skills** will be learned (Kirkpatrick & Kirkpatrick, 2006) (a rudimentary understanding of how to evaluate reactions, develop reactionnaires etc. but the skill attainment may be basic at this point. Intermediate and Advanced skills may be observed in student(s) who have had organizational experience in areas of Evaluation, Human Resources, Training and Development).

c. **Attitudes** will be minimally impacted (Kirkpatrick & Kirkpatrick, 2006).

    i. **Scope of the project**: Measurement of skills and attitudes are beyond the scope of this project.

4. The class is residential at Bloomington University i.e. only taught face-to-face.

ALIGNMENT OF KSA (Kirkpatrick & Kirkpatrick, 2006) AND REVISED BLOOM'S TAXONOMY (Forehand, 2005) IN THE DEVELOPMENT OF COGNITIVE TEST QUESTIONS FOR BLOOMINGTON UNIVERSITY



Note:
The above diagram is a depiction of KSA (Kirkpatrick & Kirkpatrick, 2006) alignment to Bloom's Taxonomy. It was important to illustrate this because the increase in the level of skill, knowledge and thus attitude will not take place at the same rate and/or magnitude for the students in R561. So while the objective for Helios is to develop test questions to determine whether students have mastered an understanding of Kirkpatrick's Evaluation Model, we conjecture that every student (as shown above Student #1 and Student #2) in the R561 course will have a unique graph due to an individual's pace of comfort in learning.

VALIDATING LEARNING

The following section outlines the learning objectives for Weeks 1 – 3 and the alignment of the objectives to the cognitive test types and Bloom's Taxonomy.

WEEK1: Overview of Kirkpatrick Model

1-1. Identify the purposes of evaluation of each of Kirkpatrick's four levels.

1-2. Discuss the strengths and the limitations of Kirkpatrick's model of evaluation.

WEEK2: Kirkpatrick Level 1: Reaction

2-1. Define the term "reactionnaire".

2-2. Describe the purpose of Kirkpatrick's model level 1 "reaction".

2-3. Describe the strengths and weaknesses of Kirkpatrick's model level 1 "reaction".

2-4. List the necessary steps to write a reactionnaire.

WEEK3: Kirkpatrick Level2: Learning

3-1. Define the "learning" in the Kirkpatrick's model.

3-2. Describe the purpose of Kirkpatrick's model level 2 "learning".

3-3. Explain how to write learning objectives.

3-4. Describe the strengths and the weaknesses of different types of test items.

3-5. Explain the concepts of reliability and validity as they relate to testing the learning.

3-6. List the necessary steps to write a written test.

LEARNING OBJECTIVES ALIGNMENT			
Learning Objective	**Test #**	**Test Item**	**Bloom's Taxonomy**
1-1	**8**	Completion	Level 1
	15	Matching	Level 1
1-2	**7**	Completion	Level 1
	16	Essay	Level 2
2-1	**13**	Multiple-Choice	Level 2
2-2	**10**	Multiple-Choice	Level 1
	12	Multiple-Choice	Level 1
2-3	**16**	Essay	Level 2
	2	True/False	Level 1
2-4	**4**	True/False	Level 1
	18	Essay	Level 1 - 3
3-1	**3**	True/False	Level 2
3-2	**11**	Multiple-Choice	Level 2
	9	Completion	Level 1
3-3	**1**	True/False	Level 1
3-4	**14**	Matching	Level 1
	17	Essay	Level 2
3-5	**6**	Completion	Level 1
	5	True/False	Level 1
3-6	**19**	Essay	Level 1

TEST DETAILS

- The test was distributed by Helios in Week 4
- Students of the class were informed that the purpose of the evaluation exercise was to assess whether the students have mastered an understanding of Kirkpatrick's evaluation model
- Paper based tests were distributed in-class by Helios representatives
- Cognitive Test Questions include a blend of completion items, multiple choice, true-false, matching and essay
  - Total questions = 19 [Completion items = 4; Multiple Choice = 4; True-False = 5; Matching = 2; Essay = 4]
- The test has **no time limits** assigned to students
  - **Note 1:** Helios believes that the purpose of the project is to conduct an interim evaluation on whether the students have mastered an understanding of Kirkpatrick's evaluation model in R561.
- The evaluation **will not depict** the points associated with the test questions

HELIOS' VALUE-ADDED APPROACH

- No time limits will be assigned to this evaluation exercise
  - The consultants will note how long it takes each student to complete the test [For e.g. if it takes Joe Smith 20 minutes to complete the questions and Jane Doe takes the entire duration of the class (i.e. 165 minutes) to complete the questions, the consultants will covertly make a note of it after each student hands in the assignment]
- 'Time' will be compared to the 'Level of Accuracy Achieved on the Test' [For instance, it took Joe Smith 20 minutes to finish the test but his level of accuracy was 60% when compared to Jane Doe whose level of accuracy was 100%]
- Each test will be followed by a brief paper-based survey (reactionnaire)
  - The survey will comprise of questions such as:
    - Do the test questions do justice evaluating your understanding of the Kirkpatrick's Model of Evaluation? Please explain your response.
    - How would you like us to evaluate your learning from this course?
    - How do you know that you have mastered a given concept?
    - What was the best part of this evaluation?
    - What were the drawbacks of this evaluation?
- The survey responses will be correlated to the other factors discussed above (i.e. time and level of accuracy).

CHANGES IMPLEMENTED

This section discusses the changes that were implemented within the test items and context.

Feedback: Test Items

Prior to disseminating the cognitive test items to the R561 class, the test questions had undergone an evaluation by Wise Giants, subject matter experts; the feedback is summarized as follows:

FEEDBACK PROVIDED	REVISIONS
GENERAL INFORMATION ON TEST MATERIALS	Reformatted the test materials (e.g. course information, instructor's name, test item numbering, etc.)
TEST TIME	Removed the test time
TEST QUESTIONS	Revised the questions (e.g. Essay question #3, T/F #4, etc.) Added test items (e.g. T/F #5, Essay question #4, etc.) Removed test items (e.g. T/F #1, Completion Item #1, etc.)
NUMBERING PROBLEM	Reformatted the numbers
GRAMMAR / WORDING	Fixed grammars and incorrect words
LEVEL OF THE TEST	N/A
NOTE: Feedback details provided in the Appendix (See page 15)	

Other Changes

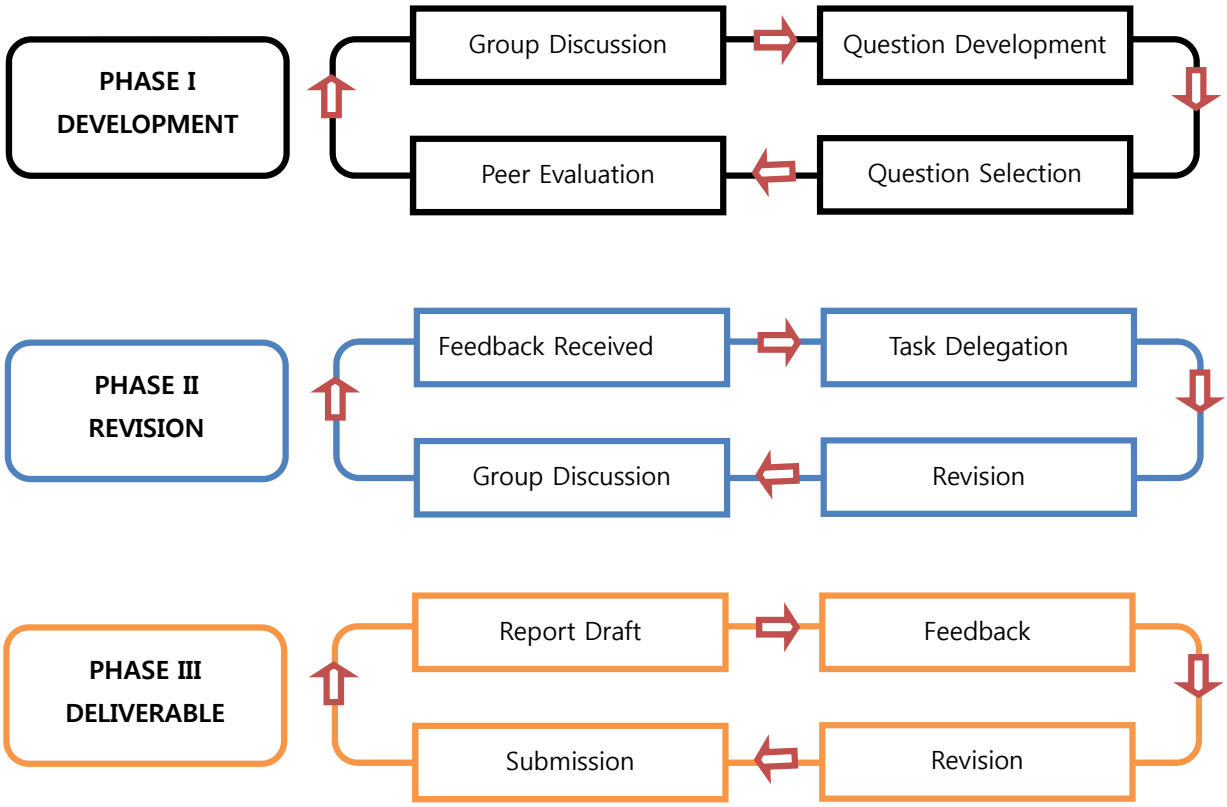
- The context was refined from the first iteration due to feedback from the reviewers that the level of questions posed was rather simple. This led us to refine and question our underlying assumption. Further additions/changes were incorporated to support the context enhancement:
  - Course outline and objectives
  - Weekly objectives and its alignment to the test questions
  - Assumptions
  - Alignment of KSA and Bloom's Taxonomy Model
  - Outline of the process adopted by Helios and concerns raised by the team
    - The concerns led to the development of recommendations and inputs to context refinement/ process improvement

SUGGESTIONS THAT WERE NOT IMPLEMENTED

- The reviewers suggested removal of T/F that was negatively worded. However, literature suggests negative words can be used if they were highlighted (Shrock & Coscarelli, 2007) and as a result we have retained the negatively worded T/F. (see Appendix A: Final Draft of Test Questions; Question 4)

PROCESS ADOPTED BY HELIOS

This section briefly outlines the process adopted by Helios:



The following section outlines the issues that were raised during each phase of the process (illustrated above). Some of which have acted as recommendations and/or as inputs for improvement:

PROCESS PHASE	ISSUES RAISED
Phase I: Development	Absence of alignment between Learning Objectives and Test Questions
	Appropriateness of Cognitive Test Questions for a Master's level course (i.e. dependence on memorization and recall)
	Alignment of KSA (Kirkpatrick & Kirkpatrick, 2006) and Bloom's Taxonomy
	Development of questions appropriate to Bloom's Level (e.g. level of difficulty) attained in Week 3 of R561
Phase II: Revision	Incorporation of feedback (decision to incorporate suggestion vs. disregard)
Phase III: Deliverable	Accurate depiction of the overall cognitive test development process

PROPOSED RECOMMENDATIONS

- ❒ Similar approach can be used on a periodic basis (also known as formative evaluation) to assess whether learning is taking place as per the objectives of R561
- ❒ Attainment of Level 4, which is the goal of R561 needs to be assessed, which could be a final project (summative evaluation) that involves synthesis of the material covered during the course
- ❒ Reduced emphasis on cognitive test questions that measure recall and memorization to assess whether students have mastered an understanding of Kirkpatrick's Model of Evaluation
    - ✍ Provide students with options (such as tests, projects, paper etc.) to evaluate their own levels of learning
- ❒ Consulting projects may be assigned to students where the primary objective may be to co-develop and assess learning (along with Training and Development) among employees within an organization
    - ✍ Exposure to the applicability of Kirkpatrick's Model of Evaluation in the real-world

Note:

- ❒ The recommendations will also be augmented from Helios' findings of correlations of time taken to complete the test, level of accuracy and the survey responses.

APPENDIX A: FINAL DRAFT OF TEST QUESTIONS

R561
EVALUATION MODULE 1

COURSE INSTRUCTOR: DR. JONES **STUDENT NAME: _____**

TRUE-FALSE
Read the following statements and then circle "True" if the statement is true and "False" if it is false:

1.	Interviews and tests are some of the methods used to measure learning.	True	False
2.	To get a 100 percent immediate response, the evaluator should ask test takers to complete the survey immediately after leaving the classroom.	True	False
3.	There is a sequence for the guidelines for evaluating learning.	True	False
4.	In a survey design, you can use both an ordinal scale and an interval scale at the same time.	True	False
5.	Validity means the degree to which measures obtained with an instrument are consistent measures of what the instrument is intended to measure.	True	False

COMPLETION ITEMS
Complete the following sentences with the correct word(s):

6. The group that does not receive the training is called the _____ group.
7. The issue with the _____ is the lack of explicit causal relationships among the different levels
8. To evaluate program the four levels represent a _____ of ways.
9. The three things that can be accomplished in learning are _____, _____, and/or _____.

MULTIPLE CHOICES
Choose the correct answer then circle the corresponding letter:

10. The purpose of evaluating reaction is to help the _____
    a. company promote the program in future
    b. trainees realize their abilities
    c. trainers increase their skills
    d. customer get more reliable service

11. Evaluation of learning is important to measure _____
    a. individual productivity
    b. individual contribution to organizational learning
    c. the effectives of the instructor in changing attitudes
    d. the impact on organizational results

12. Reaction is used to assess _____
    a. What have the participants learned.
    b. What was their contribution to the organization.
    c. Whether there was a change in behavior.
    d. Whether the training was useful.

13. According to Kirkpatrick, what **CANNOT** be considered as a reactionnaire question?
    a. A reactionnaire is to be used as a self-reporting instrument.
    b. A reactionnaire is to gather information about the expectations of the training.
    c. A reactionnaire is to measure perceived transfer of the training.
    d. A reactionnaire is to assess customer satisfaction.

MATCHING

Match the items in column A to column B. Write the numerical response (Column B) adjacent to the alphabet (Column A) in the space provided for below each table. There is only one choice for each:

14.

Column A	Column B
A. Interviews	1. Non-threatening
B. Questionnaires	2. Close relation to job performance
C. Observation	3. Low purchase cost
D. Simulation	4. Low cost to administer
E. Written Test	5. Labor Intensive
	6. Indirect and not always usable

A = ____ B = ____ C = ____ D = ____ E = ____

15.

Column A	Column B
A. Reaction	1. Measures include observation, surveys, focus groups
B. Behavior	2. Measures include pre- and post-tests, observations and role plays
C. Results	3. Measured with a happiness sheet
D. Learning	4. Measures include higher sales, fewer injuries, reduced number of errors
	5. Measures employee workload

A = ____ B = ____ C = ____ D = ____

ESSAY

Read the following questions and then answer in the space provided:

16. Briefly discuss the reason why Kirkpatrick model's first and second levels are being measured often in an organizational setting?

17. According to you, what are two main limitations for evaluating learning? Discuss.

18. How would you design a survey for customer satisfaction based on your understanding of the guidelines of Kirkpatrick Level 1: Reaction.

19. What rules should we consider when designing the questions of T/F, Completion items, Multiple-choice, Matching, and essay?

APPENDIX B: ANSWER SHEET FOR FINAL DRAFT OF TEST

ANSWER SHEET				
1. TRUE	4. FALSE	7. Kirkpatrick's model	10. a.	13. c.
2. FALSE	5. FALSE	8. sequence	11. c.	14. 5, 4, 6, 2, 1
3. TRUE	6. control	9. knowledge, skills, attitude	12. d.	15. 3, 1, 4, 2

APPENDIX C: FEEDBACK DETAILS [REVIEWER: WISE GIANTS]

QUESTION TYPES	VIOLATED DEVELOPMENTAL RULES	TEST ITEMS
Completion Items	• one single correct answer if possible	• Removed
Multiple-Choice	• avoid negative phrasing • distribute correct response order	• Negative statements used are highlighted • Number reorganized
True/False	• use slightly more false than true	• Removed one True statement and added one False statement (T: 2, F:3)
Matching	• explain if matching is 1:1 or not	• Include in the instruction
Essay	• favor more items and shorter answers over fewer items with longer answers	• Essays are reasonably short • Instructions limited the length of the answers
Note: Each reviewer from Wise Giant (Annisa, Ingu and Susan) had the same comment on the violation of developmental rules discussed in this table for all question types.		

FEEDBACK	COMMENTS	STEPS TAKEN
GENERAL INFORMATION ON TEST MATERIALS	• Need of course code/number, course number and instructor's name	• Test material design issue (FIXED)
	• Inform formative or summative evaluation exam	• N/A
TEST TIME	• Allocation of time being long (Recommendation: 22 minutes for answering T/F, completion items, multiple choice and matching. 28 minutes for answering the essay)	• Test time removed
TEST QUESTIONS	• 1-2 more essay questions	• Add one essay question
	• Essay question #3 being confusing (Recommendation: mention the context)	• Add Context: How would you design a survey for custo mer satisfaction based on your understanding of the guid eline of Kirkpatrick model's Level 1: Reaction?
	• Completion Item #3	• Removed
NUMBERING PROBLEM	• Inconsistency in putting the number	• Test material design issue (FIXED)
GRAMMAR / WORDING	• Multiple choices question 2	• Effective → Effectiveness (FIXED)
LEVEL OF THE TEST	• Test questions being easy	• The rationale on the level of the test is addressed in the context.

APPENDIX D: FIRST DRAFT OF THE CONTEXT

HELIOS, a consulting firm, is hired from the IST department at Bloomington University in order to develop a test for the course R561: Evaluation in the Instructional Development Process.

The instructor of this course, Professor Jimmy Jones, requested help to develop the test of each module to measure the students' learning as part of formative evaluation of this course. The course is designed as a 15-week class and every student is going to be tested in it every fourth week. (*Note: The final module will be tested on week 15.)

As the fundamental elements of this evaluation course lie on the Kirkpatrick's model, which is widely used by many practitioners, Professor Jones plans to have four modules to test the learning of students. The four modules will be based on the following outline:

1. Fundamental Learning: Kirkpatrick's model level 1 & 2
2. Fundamental Learning: Kirkpatrick's model level 3 & 4
3. Case Studies: Analyze and synthesize a case based on the Kirkpatrick's model
4. Case Studies: Evaluate a case based on the Kirkpatrick model

Professor Jones also strongly believes in the importance and effectiveness of the Kirkpatrick's model in this evaluation course, thus he wants to make sure that students fully understand the fundamentals of each level so that they can apply their learning to authentic cases.

The first test, Evaluation Module 1 will be taken on the fourth week of the course when students have learned the Kirkpatrick's model level 1 (Reaction) and 2 (Learning). Considering the fact that this test is the first module, Professor Jones does not intend to cover the higher levels of Bloom's Taxonomy, i.e. analysis, synthesis and evaluation, in the test. However, he wants to test whether or not the students have reached the lower levels of the taxonomy, i.e. knowledge, comprehension and application. This will help him to decide whether he can provide practical cases of implementing the Kirkpatrick's model in order for them to reach the higher levels of Bloom's Taxonomy.

In addition, Professor Jones requested the HELIOS to develop the test based on five different types of test questions: true-false, completion items, multiple-choice, matching, and essay. This is to provide students an experience of taking a well-designed test to scaffold their understanding of creating valid and reliable test items.

The Evaluation Module 1 consists of a series of questions:

1. True-False (questions)
2. Completion Items (questions)
3. Multiple-Choice (questions)
4. Matching (questions)
5. Essay (questions)

The questions were created and revised based on validity and reliability check results. We suggested Professor Jones to give at least an hour and a half (90 minutes) to complete the test.

APPENDIX E: FIRST DRAFT OF TEST QUESTIONS

EVALUATION MODULE 1

TRUE-FALSE

Read the following statements and then circle "True" if the statement is true and "False" if it is false:

1.	Interviews and test are some of the methods used to measure learning.	True	False
2.	To get a 100 percent immediate response, the evaluator should ask test takers to complete the survey immediately after leaving the classroom.	True	False
3.	The evaluators should evaluate knowledge, skills, and/or attitudes both before and after the training.	True	False
4.	There is a sequence for the guidelines for evaluating learning	True	False

COMPLETION ITEMS

Complete the following sentences with the correct word(s):

1. The group that does not receive the training is called the _____ group.
2. The issue with the _____ is the lack of explicit causal relationships among the different levels
3. As evaluators move from one level to the next, the process becomes more _____ and _____.
4. To evaluate program the four levels represent a _____ of ways

MULTIPLE CHOICES

Choose the correct answer then circle the corresponding letter:

1. The purpose of evaluating reaction is to _____
   a. help company promote the program in future
   b. help trainees realize their abilities
   c. help trainers increase their skills
   d. help customer get more reliable service

2. Evaluation of learning is important _____
   a. To measure the impact on organizational results
   b. To measure individual contribution to organization learning
   c. To measure the effectives of the instructor in changing attitudes
   d. To measure individual productivity

3. Reaction is used to assess _____
   a. What have the participants learned
   b. What was their contribution to the organization
   c. Whether the training was useful
   d. Whether there was a change in behavior
4. According to Kirkpatrick, what can NOT be considered as a reactionnaire question?
   e. A reactionnaire is to use as a self-reporting instrument
   f. A reactionnaire is to gather information about expectations of the training
   g. A reactionnaire is to measure perceived transfer of the training
   h. A reactionnaire is to assess customer satisfaction

MATCHING

Match the items in column A to column B. Write the numerical response (Column B) adjacent to the alphabet (Column A) in the space provided for below each table:

I.

Column A	Column B
F. Interviews	7. Non-threatening
G. Questionnaires	8. Close relation to job performance
H. Observation	9. Low purchase cost
I. Simulation	10. Low cost to administer
J. Written Test	11. Labor Intensive
	12. Indirect and not always usable

A = ____ B = ____ C = ____ D = ____ E = ____

II.

Column A	Column B
E. Reaction	6. Measures include observation, surveys, focus groups
F. Behavior	7. Measures include pre and post tests, observations and role plays
G. Results	8. Measured with a happiness sheet
H. Learning	9. Measures include higher sales, fewer injuries, reduced number of errors
	10. Measures employee workload

A = ____ B = ____ C = ____ D = ____

ESSAY

Read the following questions and then answer in the space provided:

Briefly discuss the reason why Kirkpatrick model's first and second levels are being measured often in an organizational setting?

According to you, what are the limitations for evaluating learning? Discuss.

Discuss the steps that you would undertake based on your understanding of the guidelines of Kirkpatrick Level 1.

APPENDIX F: INDIVIDUAL REVIEWER ASSESSMENTS

This section has three individual assessments on the first draft of the Helios' Test Development.

Assessing Written (Cognitive) Test Questions

Developer Name (Helios) Reviewer Name (Wise Giants)

Question Types	Development Rules	Checklists
Completion Items	• omit only significant words	Yes (√) No ( ) NA ( )
	• one single correct answer if possible	Yes ( ) No (√) NA ( )
	• use blanks of the same length	Yes (√) No ( ) NA ( )
	• avoid grammatical clues	Yes (√) No ( ) NA ( )
	• place blanks at the end if possible	Yes (√) No ( ) NA ( )
	• avoid direct quotes from the text	Yes (√) No ( ) NA ( )
Multiple-Choice	• use direct, explicit questions	Yes (√) No ( ) NA ( )
	• ask only one thing	Yes (√) No ( ) NA ( )
	• no excess or irrelevant info	Yes (√) No ( ) NA ( )
	• phrase to avoid repetition of words	Yes (√) No ( ) NA ( )
	• avoid negative phrasing	Yes ( ) No (√) NA ( )
	• make all choices plausible	Yes (√) No ( ) NA ( )
	• mutually exclusive choices	Yes (√) No ( ) NA ( )
	• arrange choices logically	Yes (√) No ( ) NA ( )
	• one clearly best response	Yes (√) No ( ) NA ( )
	• choices equal in length and phrasing	Yes (√) No ( ) NA ( )
	• No irrelevant clues via language	Yes (√) No ( ) NA ( )
	• at least four alternatives each	Yes (√) No ( ) NA ( )
	• distribute correct response order	Yes ( ) No (√) NA ( )
	• Use "none" and "all" sparingly	Yes (√) No ( ) NA ( )
True-False	• Express the item simply and clearly	Yes (√) No ( ) NA ( )
	• Express a single idea in each item	Yes (√) No ( ) NA ( )
	• Include sufficient background info	Yes (√) No ( ) NA ( )
	• Avoid direct quotes from readings	Yes (√) No ( ) NA ( )
	• Phrase statements positively	Yes (√) No ( ) NA ( )
	• Avoid unfamiliar vocabulary	Yes (√) No ( ) NA ( )
	• Avoid determiners like "all" which permit test gamesmanship	Yes (√) No ( ) NA ( )
	• use slightly more false than true	Yes ( ) No (√) NA ( )
Matching	• clearly state basis for matching	Yes (√) No ( ) NA ( )
	• explain if matching is 1:1 or not	Yes ( ) No (√) NA ( )
	• indicate where to write answer	Yes (√) No ( ) NA ( )
	• use only homogeneous material	Yes (√) No ( ) NA ( )
	• put responses in systematic order	Yes (√) No ( ) NA ( )
	• avoid grammatical or other clues	Yes (√) No ( ) NA ( )
	• use 10 or fewer stimuli	Yes (√) No ( ) NA ( )
	• use more responses than stimuli	Yes (√) No ( ) NA ( )
	• use short phrases or single words	Yes (√) No ( ) NA ( )
Essay	• clearly define task	Yes (√) No ( ) NA ( )
	• Favor more items and shorter answers over fewer items with longer answers	Yes ( ) No (√) NA ( )
	• Avoid using optional questions	Yes (√) No ( ) NA ( )
Overall Comments	✓ The layout is good and easy to read (not too crowded). ✓ It would be better if you put the course code/number, course number and instructor's name on the question sheet. ✓ The time allocation seems too long for answering all these questions. In my opinion, 50-60 minutes is adequate to finish this test (22 minutes for answering T/F, completion items, multiple choice and matching. 28 minutes for answering the essay). ✓ In my opinion, essay question no.3 is confusing. You need to mention the context. ✓ There is inconsistency in putting the number for essay and matching questions with the previous	

Assessing Written (Cognitive) Test Questions

Developer Name (Helios) Reviewer Name (Wise Giants)

	questions (T/F, completion items, and multiple choices). ✓ In multiple choices question no.2 option c, is it effectives or effectiveness?

Assessing Written (Cognitive) Test Questions

Developer Name (Helios) Reviewer Name (Wise Giants)

Question Types	Development Rules	Checklists
Completion Items	• omit only significant words	Yes (√) No ( ) NA ( )
	• one single correct answer if possible	Yes ( ) No (√) NA ( )
	• use blanks of the same length	Yes (√) No ( ) NA ( )
	• avoid grammatical clues	Yes (√) No ( ) NA ( )
	• place blanks at the end if possible	Yes (√) No ( ) NA ( )
	• avoid direct quotes from the text	Yes (√) No ( ) NA ( )
Multiple-Choice	• use direct, explicit questions	Yes (√) No ( ) NA ( )
	• ask only one thing	Yes (√) No ( ) NA ( )
	• no excess or irrelevant info	Yes (√) No ( ) NA ( )
	• phrase to avoid repetition of words	Yes (√) No ( ) NA ( )
	• avoid negative phrasing	Yes ( ) No (√) NA ( )
	• make all choices plausible	Yes (√) No ( ) NA ( )
	• mutually exclusive choices	Yes (√) No ( ) NA ( )
	• arrange choices logically	Yes (√) No ( ) NA ( )
	• one clearly best response	Yes (√) No ( ) NA ( )
	• choices equal in length and phrasing	Yes (√) No ( ) NA ( )
	• No irrelevant clues via language	Yes (√) No ( ) NA ( )
	• at least four alternatives each	Yes (√) No ( ) NA ( )
	• distribute correct response order	Yes ( ) No (√) NA ( )
	• Use "none" and "all" sparingly	Yes (√) No ( ) NA ( )
True-False	• Express the item simply and clearly	Yes (√) No ( ) NA ( )
	• Express a single idea in each item	Yes (√) No ( ) NA ( )
	• Include sufficient background info	Yes (√) No ( ) NA ( )
	• Avoid direct quotes from readings	Yes (√) No ( ) NA ( )
	• Phrase statements positively	Yes (√) No ( ) NA ( )
	• Avoid unfamiliar vocabulary	Yes (√) No ( ) NA ( )
	• Avoid determiners like "all" which permit test gamesmanship	Yes (√) No ( ) NA ( )
	• use slightly more false than true	Yes ( ) No (√) NA ( )
Matching	• clearly state basis for matching	Yes (√) No ( ) NA ( )
	• explain if matching is 1:1 or not	Yes ( ) No (√) NA ( )
	• indicate where to write answer	Yes (√) No ( ) NA ( )
	• use only homogeneous material	Yes (√) No ( ) NA ( )
	• put responses in systematic order	Yes (√) No ( ) NA ( )
	• avoid grammatical or other clues	Yes (√) No ( ) NA ( )
	• use 10 or fewer stimuli	Yes (√) No ( ) NA ( )
	• use more responses than stimuli	Yes (√) No ( ) NA ( )
	• use short phrases or single words	Yes (√) No ( ) NA ( )
Essay	• clearly define task	Yes (√) No ( ) NA ( )
	• Favor more items and shorter answers over fewer items with longer answers	Yes ( ) No (√) NA ( )
	• Avoid using optional questions	Yes (√) No ( ) NA ( )
Overall Comments	Comparing to total test time, 90 minutes, the number of test questions is much few. More test questions should be developed. Overall, test questions seem to be easy. The level of difficulty of test questions are considered in order to identify the degree of achievement of students Your team needs to include basic information about the course, such as the course name and formative or summative evaluation exam within the part of the beginning or ending.	

Assessing Written (Cognitive) Test Questions

Developer Name (Helios) Reviewer Name (Wise Giants)

Question Types	Development Rules	Checklists
Completion Items	• omit only significant words	Yes (√) No ( ) NA ( )
	• one single correct answer if possible	Yes ( ) No (√) NA ( )
	• use blanks of the same length	Yes (√) No ( ) NA ( )
	• avoid grammatical clues	Yes (√) No ( ) NA ( )
	• place blanks at the end if possible	Yes (√) No ( ) NA ( )
	• avoid direct quotes from the text	Yes (√) No ( ) NA ( )
Multiple-Choice	• use direct, explicit questions	Yes (√) No ( ) NA ( )
	• ask only one thing	Yes (√) No ( ) NA ( )
	• no excess or irrelevant info	Yes (√) No ( ) NA ( )
	• phrase to avoid repetition of words	Yes (√) No ( ) NA ( )
	• avoid negative phrasing	Yes ( ) No (√) NA ( )
	• make all choices plausible	Yes (√) No ( ) NA ( )
	• mutually exclusive choices	Yes (√) No ( ) NA ( )
	• arrange choices logically	Yes (√) No ( ) NA ( )
	• one clearly best response	Yes (√) No ( ) NA ( )
	• choices equal in length and phrasing	Yes (√) No ( ) NA ( )
	• No irrelevant clues via language	Yes (√) No ( ) NA ( )
	• at least four alternatives each	Yes (√) No ( ) NA ( )
	• distribute correct response order	Yes ( ) No (√) NA ( )
	• Use "none" and "all" sparingly	Yes (√) No ( ) NA ( )
True-False	• Express the item simply and clearly	Yes (√) No ( ) NA ( )
	• Express a single idea in each item	Yes (√) No ( ) NA ( )
	• Include suffcient background info	Yes (√) No ( ) NA ( )
	• Avoid direct quotes from readings	Yes (√) No ( ) NA ( )
	• Phrase statements positively	Yes (√) No ( ) NA ( )
	• Avoid unfamiliar vocabulary	Yes (√) No ( ) NA ( )
	• Avoid determiners like "all" which permit test gamesmanship	Yes (√) No ( ) NA ( )
	• use slightly more false than true	Yes ( ) No (√) NA ( )
Matching	• clearly state basis for matching	Yes (√) No ( ) NA ( )
	• explain if matching is 1:1 or not	Yes ( ) No (√) NA ( )
	• indicate where to write answer	Yes (√) No ( ) NA ( )
	• use only homogeneous material	Yes (√) No ( ) NA ( )
	• put responses in systematic order	Yes (√) No ( ) NA ( )
	• avoid grammatical or other clues	Yes (√) No ( ) NA ( )
	• use 10 or fewer stimuli	Yes (√) No ( ) NA ( )
	• use more responses than stimuli	Yes (√) No ( ) NA ( )
	• use short phrases or single words	Yes (√) No ( ) NA ( )
Essay	• clearly define task	Yes (√) No ( ) NA ( )
	• Favor more items and shorter answers over fewer items with longer answers	Yes ( ) No (√) NA ( )
	• Avoid using optional questions	Yes (√) No ( ) NA ( )
Overall Comments	✧ Outline is clear and well-organized. ✧ Completion Item 3: I think there could be many possible answers other than two. ✧ Multiple-Choice 4: avoid negative phrasing ✧ It would be better if you provide some space that a test-taker can write his/her name. ✧ I think you can make 1-2 more essay questions, if possible. The number of questions seems not enough for the total test time, 90 min.	

APPENDIX G: DETAILED ALIGNMENT OF LEARNING OBJECTIVE, TEST ITEMS AND BLOOM'S TAXONOMY

OBJECTIVE	TEST ITEM #	BLOOM'S TAXONOMY	TEST ITEM	QUESTION	
1-1	8	Remember	Completion	To evaluate program the four levels represent a _____ of ways.	
	15	Remember	Matching	**Column A** A. Reaction B. Behavior C. Results D. Learning	**Column B** 1. Measures include observation, surveys, focus groups 2. Measures include pre and post tests, observations and role plays 3. Measured with a happiness sheet 4. Measures include higher sales, fewer injuries, reduced number of errors 5. Measures employee workload
1-2	7	Remember	Completion	The issue with the _____ is the lack of explicit causal relationships among the different levels.	
	16	Understanding	Essay	Briefly discuss the reason why Kirkpatrick model's first and second levels are being measured often in an organizational setting?	
2-1	13	Understanding	Multiple-Choice	According to Kirkpatrick, what can NOT be considered as a reactionnaire question? a. A reactionnaire is to be used as a self-reporting instrument b. A reactionnaire is to gather information about the expectations of the training c. A reactionnaire is to measure perceived transfer of the training d. A reactionnaire is to assess customer satisfaction	
2-2	10	Remember	Multiple-Choice	The purpose of evaluating reaction is to help the: a. company promote the program in future b. trainees realize their abilities c. trainers increase their skills d. customer get more reliable service	
	12	Remember	Multiple-Choice	Reaction is used to assess: a. What have the participants learned b. What was their contribution to the organization c. Whether the training was useful d. Whether there was a change in behavior	
2-3	16	Understanding	Essay	Briefly discuss the reason why Kirkpatrick model's first and second levels are being measured often in an organizational setting?	
	2	Remember	T/F	To get a 100 percent immediate response, the evaluator should ask test takers to complete the survey immediately after leaving the classroom.	
2-4	4	Remember	T/F	In a survey design, you CANNOT use both an ordinal scale and an interval scale at the same time.	
	18	Remember Understanding Applying	Essay	How would you design a survey for customer satisfactio n based on your understanding of the guidelines of Kirkpatrick Level 1: Reaction	
3-1	3	Understanding	T/F	There is a sequence for evaluating the guidelines for learning.	
3-2	11	Understanding	Multiple-Choice	Evaluation of learning is important to measure: a. the impact on organizational results b. individual contribution to organization learning c. the effectives of the instructor in changing attitudes d. individual productivity	
	9	Remember	Completion	The three things that can be accomplished in learning a re _____, _____, and/or _____.	
3-3	1	Remember	T/F	Interviews and tests are some of the methods used to measure learning.	

3-4	**14**	**Remember**	**Matching**	**Column A** A. Interviews B. Questionnaires C. Observation D. Simulation E. Written Test	**Column B** 1. Non-threatening 2. Close relation to job performance 3. Low purchase cost 4. Low cost to administer 5. Labor intensive 6. Indirect and not always usable.
	17	**Understanding**	**Essay**	According to you, what are two main limitations for evaluating learning? Discuss.	
3-5	**6**	**Remember**	**Completion**	The group that does not receive the training is called the _____ group.	
	5	**Remember**	**T/F**	Validity means the degree to which measures obtained with an instrument are consistent measures of what the instrument is intended to measure.	
3-6	**19**	**Remember**	**Essay**	What rules should we consider when designing the questions of T/F, Completion items, Multiple-choice, Matching, and essay?	

REFERENCES

Forehand, M. (2005). *Terminology Changes*. Retrieved February 28, 2010, from Bloom's Taxonomy: http://projects.coe.uga.edu/epltt/index.php?title=Bloom%27s_Taxonomy#end

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating Training Programs.* San Francisco: Berrett-Koehler Publishers, Inc.

Shrock, S., & Coscarelli, W. (2007). *Criterion-Referenced Test Development.* San Francisco: Pfeiffer.