Will Thalheimer, PhD

The Learning-Transfer Evaluation Model (LTEM) and Fundamental Secrets in Learning Evaluation

1. Do you have a philosophy or approach to learning evaluation that you embrace? If so, what is it?

2. What’s the best reasoning you’ve heard for why we should evaluate learning?
Agenda

- Fundamentals of Learning Evaluation
- A Long History of Heartbreak and Disillusionment
- Biases, Mistakes, and Some Perspective
- LTEM – Learning-Transfer Evaluation Model
- Tier 3A – Performance-Focused Smile Sheet Questions
- The SEDA Model
- Tier 5 – Scenario-Based Questions
- Practice and Feedback

Slides available at: www.is.gd/will999stuff
<table>
<thead>
<tr>
<th>The Decisive Dozen for Learning Design and Learning Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Content</td>
</tr>
<tr>
<td>2. Exposure</td>
</tr>
<tr>
<td>3. Guiding Attention</td>
</tr>
<tr>
<td>4. Creating Correct Conceptions</td>
</tr>
<tr>
<td>5. Repetition</td>
</tr>
<tr>
<td>6. Feedback</td>
</tr>
<tr>
<td>7. Variation</td>
</tr>
<tr>
<td>8. Retrieval Practice</td>
</tr>
<tr>
<td>9. Context Alignment</td>
</tr>
<tr>
<td>10. Spacing</td>
</tr>
<tr>
<td>11. Persuasion</td>
</tr>
<tr>
<td>12. Perseverance</td>
</tr>
</tbody>
</table>

http://is.gd/ddResearch
Quite simply, the BEST book on smile sheet creation and utilization, Period!

Karl M. Kapp
Professor of Instructional Technology
Bloomsburg University

Thoughtful and sensible advice for feedback tools that will provide valid and actionable data.

Robert O. Brinkerhoff
Professor Emeritus, Western Michigan University
& Director, Brinkerhoff Evaluation Institute

Evidence-based practice at the master level.

Julie Dirksen
Author of Design For How People Learn
Why is this farmer gathering and analyzing growing rates, crop yields, infestation rates, soil quality, amount of fertilizer, rainfall, sunshine, revenue per sales channel, et cetera?
Data & Analysis
What criteria should he have for the data he’s collecting?

• It should be ACCURATE/VALID
• It should be RELEVANT
• It should be HIGHLY PREDICTIVE
• It should be COST EFFECTIVE
• Most importantly, it should enable his organization to make its most important DECISIONS!

Data & Analysis
How are we doing in collecting data that:

• Is ACCURATE/VALID
• Is RELEVANT
• Is HIGHLY PREDICTIVE
• Is COST EFFECTIVE
• Helps us make our most important DECISIONS!

Learning Professionals
### Data & Analysis

<table>
<thead>
<tr>
<th>How are we doing in collecting data that:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Is ACCURATE/VALID</td>
</tr>
<tr>
<td>• Is RELEVANT</td>
</tr>
<tr>
<td>• Is HIGHLY PREDICTIVE</td>
</tr>
<tr>
<td>• Is COST EFFECTIVE</td>
</tr>
<tr>
<td>• Helps us make our most important DECISIONS!</td>
</tr>
</tbody>
</table>

### Learning Professionals

<table>
<thead>
<tr>
<th>What are our most important decisions?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are these helping?</td>
</tr>
<tr>
<td>What about our smile-sheet (learner-feedback) data?</td>
</tr>
<tr>
<td>What about our quiz and knowledge-check data?</td>
</tr>
<tr>
<td>What about our post-training on-the-job performance data?</td>
</tr>
</tbody>
</table>

### Will’s Learning-Evaluation Philosophy

1. Our evaluation approaches must be aligned with human learning and cognition.
2. As learning professionals we have a responsibility to improve our performance.
3. We have a responsibility to gather data that enables us to continuously improve the learning we create and deploy.
4. We need to craft our evaluation frameworks to help us make our most important decisions.
5. We have to engage in evaluation at a reasonable benefit/cost ratio.
6. To sustain our ability to create great learning-and-performance results, we have to maintain credibility with key stakeholders.
Warning!

We must look into the shadows first before we can emerge into the light

“The evaluation of training programs in terms of ‘results’ is progressing at a very slow rate.”

Donald Kirkpatrick
1960
“For the most part, the benefits of industrial training have been taken on faith. Few demands have been made to evaluate it in a rigorous manner.”

Ronald Burke
Researcher
1969

“With some notable exceptions... relatively little work has been devoted to making evaluation more useful and worthwhile.”

Robert Brinkerhoff
1981
“ASTD’s research revealed that the actual practice of evaluation doesn’t often follow the strict recommendations of evaluation literature.”

1990

Only 20% were able to do the learning measurement they wanted to do...

2007
“In every year [from 2005-2007], more than 90 percent rated measurement as the number one or number two area they would like to improve.”

Josh Bersin
Principle and Founder
Bersin, Deloitte Consulting LLP
2008

In general, are you able to do the learning measurement you want to?

Research conducted by Will Thalheimer, PhD of Work-Learning Research, Inc.
Measurement doesn’t illuminate everything. Measurement illuminates some things.

“Our measures are not perfect, but they should instead be thought of as approximations.”
Deborah L. Bandalos,
Author of the 2018 book: 
*Measurement Theory and Applications for the Social Sciences*

There is no perfect measurement tool!
Acceptable Error

Unacceptable Error

Overarching Goal of Measurement:
Being Accurate Reducing Error

When along a single dimension, key is to be close to true level.

True Speed

+ .0004

No Error

When along multiple dimensions, key is to choose correct dimensions, and the priorities between those dimensions.

Example: Manufacturing Line Performance

Measures: speed, quality, attendance, following rules, getting along

Measurement has Error

Measurements made by a thermometer are imperfect (atmospheric variables, sunlight, quality issues, etc.)

Minimizing Error

We should, ideally, take lots of measurements using different, high quality thermometers.

Multiple Measures Reduce Error

| 43.1 | 42.8 | 42.7 | 42.4 | 42.4 | 42.8 | 42.6 | 43.0 |
“Metric fixation is in fact often counterproductive, with costs to individual satisfaction with work, organizational effectiveness, and economic growth.”

Jerry Muller, in Interview with Princeton University Press

Let us be humble, skeptical, and wise!!
Hendrick had been CLO for three months and had begun to wrest control of the thrashing monster that was his company’s learning-and-development department. He started by getting learning evaluation under control, based on the dictum “what gets measured, gets managed.”

The first pilot was with the strategically-important Leadership for Hippos course. The new evaluation strategy involved the following:

- Utilizing psychometrically-validated questions on a representative sample of the highest-priority content knowledge taught in the course and divided into ten competency categories.
- Requiring employees to be correct on eight of the ten questions for each competency on a test given during the last two hours of the four-day workshop.

What do you think Sandra, a world-class learning evaluation expert, will tell Hendrick about the effectiveness of his evaluation design?

A. NOT AT ALL effective.
B. SOMEWHAT effective.
C. MOSTLY effective.
D. VERY effective.

Most importantly, what do you think are the strengths and weaknesses of Hendrik’s approach?
Remembering

Learning

Intervention

Performance

Situation

Learning

Outcomes

On-the-Job Performance

Individual

Results

Organizational

Results

The Learning Landscape
What does an end-of-course assessment tell us?

During Learning

After Learning

Learning and Forgetting Curves

© Copyright by Work-Learning Research, Inc. (www.worklearning.com)
WHEN did you Measure Learning?

WHERE did you Measure Learning?
"Level C represents the last level of certification that can be considered to assess an ability to perform on the job. Level D represents the first quantum jump away from fidelity in assessment and should be used with caution."
Alena wants to start a firm that helps farmers grow food organically. She has a degree in sustainable agriculture and has worked for a non-profit organization for seven years doing similar work. She’s developed a marketing plan, a financial plan, and has found several farmers who would pay her if she went out on her own. What should Alena do first—before she tells her boss that she’s quitting to start her own firm?

A. Create a cash flow statement to determine whether her predicted income will support the business through the first year.
B. Form a group of advisors with experience in both small-business management and agriculture.
C. Determine whether she has enough seed money and start-up capital to get started.
D. Analyze her values and goals to ensure that the proposed business will support them.

Three Biases in the Way We Measure Level 2 Learning

1. Measuring only at end of learning
2. Measuring in the learning context
3. Utilizing knowledge questions
Hendrick had been CLO for three months and had begun to wrest control of the thrashing monster that was his company’s learning-and-development department. He started by getting learning evaluation under control, based on the dictum “what gets measured, gets managed.”

The first pilot was with the strategically-important Leadership for Hippos course. The new evaluation strategy involved the following:

- Utilizing psychometrically-validated questions on a representative sample of the highest-priority content knowledge taught in the course and divided into ten competency categories.

- Requiring employees to be correct on eight of the ten knowledge questions for each competency on a test given during the last two hours of the four-day workshop.

Which of the three biases did Hendrick fail to overcome?

---

**Percent of Companies Using “To Any Extent”**

<table>
<thead>
<tr>
<th>ATD and i4cp</th>
<th>Classroom</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Behavior</strong></td>
<td>15%</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td>10%</td>
<td>14%</td>
</tr>
<tr>
<td><strong>ROI</strong></td>
<td>6%</td>
<td>6%</td>
</tr>
</tbody>
</table>

**Technology – Reaction**: 52%  
**Classroom – Reaction**: 81%

**Technology – Learning**: 43%  
**Classroom – Learning**: 50%

**Behavior**: 15%  
**Results**: 10%  
**ROI**: 6%
“The Kirkpatrick framework has a number of theoretical and practical shortcomings.”

“[It] is antithetical to nearly 40 years of research on human learning, leads to a checklist approach to evaluation (e.g., ‘we are measuring Levels 1 and 2, so we need to measure Level 3’), and, by ignoring the actual purpose for evaluation, risks providing no information of value to stakeholders…” (p. 91)
The Kirkpatrick Four-Level Model

Level 1
Reaction

Level 2
Learning

Level 3
Behavior

Level 4
Results

Read about Raymond Katzell's role:
https://is.gd/originator

The Kirkpatrick-Katzell Four-Level Model

Level 1
Reaction

Level 2
Learning

Level 3
Behavior

Level 4
Results
Common Measurement Problems

Now?
### The Learning-Transfer Evaluation Model

**1. Attitudes**  
Learner engages in activities related to learning;  
- **Skepticism of Attitudes**  
- **Inflow of Attitudes**  
- **Persistence of Attitudes**  
- **Innovations of Attitudes**  
  - Males: numerical or symbolic learning questions  
  - Females: numerical or symbolic learning questions  
  - Response: learning experiences  
  - Translates to life work  
  - Not to learn

**2. Knowledge**  
Learner answers questions about early demonstrations;  
- **Knowledge Retrieval**  
- **Knowledge Anticipation**  
- **Knowledge Learning**  
- **Knowledge Approximation**  
  - Not to learn

**3. Perceptions**  
A learner's perception is a way that meets each related to learning effectiveness;  
- **Skepticism of Perceptions**  
- **Skepticism of Learning**  
- **Inflow of Perceptions**  
- **Innovations of Perceptions**  
  - Males: numerical or symbolic learning questions  
  - Females: numerical or symbolic learning questions  
  - Response: learning experiences  
  - Not to learn

**4. Task Competence**  
Learner performs relevant realistic actions and decisions-making;  
- **Task Competence**  
- **Knowledge Anticipation**  
- **Knowledge Learning**  
- **Knowledge Approximation**

**5. Transfer**  
When learner sees what we intend to perform tasks successfully;  
- **Active Transfer**  
- **Passive Transfer**

**6. Decision-Making Competence**  
Learner makes decisions given relevant realistic situations;  
- **Decision-Making Competence**

**7. Evaluation**

---

**Special Thanks:**
- Julie Dirksen
- Clark Quinn
- Roy Pollock
- Adam Neaman
- Yvon Dalat
- Emma Weber
- Scott Weersing
- Mark Jenkins
- Ingrid Guerra-Lopez
- Rob Brinkerhoff
- Trudy Mandeville
- Mike Rustici
Gather Blind-Spot Feedback

Generate Quick Solution

Triggering Event

Grok Problem

Industry Empathy

Reality-Test Solution

Improve Solution

Generate Solution

Learning Technologies
31 January - 1 February 2018
Olympia London
What Messages Does the Four-Level Model Send?

The Kirkpatrick-Katzell Four-Level Model

Level 1: Reaction
Level 2: Learning
Level 3: Behavior
Level 4: Results
### Messaging of Four-Level Model of Learning Evaluation

<table>
<thead>
<tr>
<th>Beneficial Messages</th>
<th>Harmful Messages (Sent or Missed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Don’t Just Focus on Learning! Focus on Results too!</td>
<td>• Does Not Warn Us Against Ineffective Evaluation Practices</td>
</tr>
<tr>
<td>• Learner Opinions Are Not Most Important</td>
<td>• Ignores the Role of Remembering</td>
</tr>
<tr>
<td></td>
<td>• Level 2 Learning is Mashed into One Bucket</td>
</tr>
</tbody>
</table>

### Messaging our Evaluation Model Should Have: A Few Examples

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Just because learners ENGAGE IN LEARNING doesn’t mean they will have learned. Therefore, measuring attendance is an inadequate way of evaluating learning.</td>
</tr>
<tr>
<td>• Just because learners PAY ATTENTION doesn’t mean they learned. Measuring attention is an inadequate way of evaluating learning.</td>
</tr>
<tr>
<td>• Just because learners ACTIVELY PARTICIPATE in learning doesn’t mean they learned. Measuring participation is inadequate</td>
</tr>
<tr>
<td>• Just because learners say they LIKE A LEARNING EVENT doesn’t mean they learned. Therefore, surveying learners on their general satisfaction is an inadequate way of evaluating learning.</td>
</tr>
<tr>
<td>• Just because learners REPORT THEY HAVE EXPERIENCED EFFECTIVE LEARNING METHODS doesn’t guarantee they learned. Therefore, surveying learners on their experience with learning methods, must be augmented with objective measures of learning.</td>
</tr>
<tr>
<td>• Just because learners CAN RECITE FACTS AND TERMINOLOGY doesn’t mean they know what to do. Therefore, measuring knowledge recitation is an inadequate way of evaluating learning.</td>
</tr>
<tr>
<td>• Just because learners COMPREHEND A CONCEPT doesn’t mean they will be able to use that concept in a work situation. Therefore, measuring knowledge retention is an inadequate way of evaluating learning.</td>
</tr>
<tr>
<td>• Just because learners DEMONSTRATE COMPETENCY during a learning event doesn’t mean they’ll remember how to use the competency later. Therefore, measuring competency during or soon after a learning event is an inadequate way of evaluating learning.</td>
</tr>
<tr>
<td>• There are a NUMBER OF GOALS WE SHOULD HAVE as learning designers, including supporting our learners in building: comprehension, remembering, decision making competence, task competence, and perseverance in applying what they've learned to their job or other performance situations.</td>
</tr>
<tr>
<td>Tier</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>
| 7    | Transfer | When learner uses what was learned to perform work tasks successfully—as clearly demonstrated through objective measures.  
- Assisted Transfer—when performance is substantially prompted/supported.  
  ADEQUATE TO CERTIFY ASSISTED TRANSFER.  
- Full Transfer—when learner demonstrates full agency in applying the learning.  
  ADEQUATE TO CERTIFY FULL TRANSFER. |
| 6    | Task Competence | Learner performs relevant realistic actions and decision making.  
- Task Competence—during or right after learning event.  
  Not a fully adequate metric because learners may forget their task competencies.  
- Remembered Task Competence—after several days or more.  
  ADEQUATE TO CERTIFY TASK COMPETENCE.  
  NOTE: “Tasks” comprise both decision making and action taking. For example, a person learning to write poetry could decide to use metaphor, could act to use it, or could do both. |
| 5    | Decision Making Competence | Learner makes decisions given relevant realistic scenarios.  
- Decision Making Competence—during or right after learning event.  
  Not a fully adequate metric because learners may forget decision making competencies.  
- Remembered Decision Making Competence—after several days or more.  
  ADEQUATE TO CERTIFY DECISION MAKING COMPETENCE. |

| 4    | Knowledge | Learner answers questions about facts/terminology.  
- Knowledge Recitation—during or right after learning event.  
  Usually inadequate because testing terminology does not fully enable performance.  
- Knowledge Retention—after several days or more.  
  Usually inadequate because remembering terminology does not fully enable performance. |
| 3    | Learner Perceptions | A. Learner is queried in a way that reveals insights related to learning effectiveness.  
- Examples: Measures that target Learner Comprehension, Realistic Practice, Learner Motivation to Apply, After-Learning Support, etc.  
  Such measures can hint at outcomes but should be augmented with objective outcome measures.  
B. Learner is queried in a way that does NOT reveal insights on learning effectiveness.  
- Examples: Measures that target Learner Satisfaction, Course Reputation, etc.  
  A metric inadequate to validate learning success—because such perceptions are not always related to learning results. |
| 2    | Activity | Learner engages in activities related to learning.  
- Measures of Attention  
  A metric inadequate to validate learning success—because learners may pay attention but not learn.  
- Measures of Interest  
  A metric inadequate to validate learning success—because learners may show interest but not learn.  
- Measures of Participation  
  A metric inadequate to validate learning success—because learners may participate but not learn. |
| 1    | Attendance | Learner signs up, starts, attends, or completes a learning experience.  
- A metric inadequate to validate learning success—because learners may attend but not learn. |
Comparing the Kirkpatrick-Katzell Four-Level Model of Learning Evaluation

**Level 1 -- Reaction**
- **Learner**:
  - **Attendance**
  - **Learning Engagement**: involves reaction to the learning environment, such as attitudes, interest, and learning satisfaction.

**Level 2 -- Learning**
- **Learner**:
  - **Knowledge**: knowledge application demonstrated in the study, learning, and testing periods.
  - **Decision Making Competence**: decision-making relevance, after several days or more.

**Level 3 -- Behavior**
- **Learner**:
  - **Task Competence**: relevant realistic actions and decision making.
  - **Remembered Task Competence**: after several days or more.

**Level 4 -- Results**
- **Learner**:
  - **Assisted Transfer**: when performance is substantially prompted/supported.
  - **Full Transfer**: when learner demonstrates full agency in applying the learning.

**Work Performance**

**In Learning**

![Image of the Learning-Transfer Evaluation Model](image)

**The Learning-Transfer Evaluation Model**

---

**Certifying Effects of Transfer Requires**: Certification of transfer plus a rigorous method of assessing transfer’s causal impact—including positive and negative effects.
Two Ways to Use LTEM

1. Assessing Your Evaluations
2. Learning Design & Development: Working Backward from Your Goals

The Learning-Transfer Evaluation Model

<table>
<thead>
<tr>
<th>Year</th>
<th>Effects of Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>Sales Increase by 5%</td>
</tr>
<tr>
<td>2019</td>
<td>Managers Coach Better</td>
</tr>
<tr>
<td>2018</td>
<td>Simulated Coaching</td>
</tr>
<tr>
<td>2017</td>
<td>Scenario Questions</td>
</tr>
<tr>
<td>2016</td>
<td>IF-THEN Decisions</td>
</tr>
<tr>
<td>2015</td>
<td>Perf-Focused Questions</td>
</tr>
</tbody>
</table>

1. Sales Increase by 5%
2. Managers Coach Better
3. Simulated Coaching
4. Scenario Questions
5. IF-THEN Decisions
6. Perf-Focused Questions
There is no perfect measurement tool!

LTEM is not a panacea!

Do your evaluations with wisdom!!
Quite simply, the BEST book on smile sheet creation and utilization, Period!

Karl M. Kapp
Professor of Instructional Technology
Bloomsburg University

Thoughtful and sensible advice for feedback tools that will provide valid and actionable data.

Robert O. Brinkerhoff
Professor Emeritus, Western Michigan University & Director, Brinkerhoff Evaluation Institute

Evidence-based practice at the master level.

Julie Dirksen
Author of Design For How People Learn

How much are smile-sheet results correlated with learning outcomes?

A. High marks indicate that the training was likely to be at least SOMEWHAT SUCCESSFUL in creating learning.

B. High marks on smile sheets tell us ALMOST NOTHING about the success of our training programs in creating learning.
Trevor was recently promoted into a new position, the Director of Learning Evaluation for an integrated health network. After a month of talking with key stakeholders and examining the past few years of evaluation results, Trevor gathers his team for strategic planning.

After several weeks of researching options, Trevor’s team decides to improve their evaluations by:

• Reframing all their survey questions to be learner-centric (stated from learner’s viewpoint).
• Focusing particularly on how likely learners are to recommend the learning to others.
Trevor was recently promoted into a new position, the Director of Learning Evaluation for an integrated health network. After a month of talking with key stakeholders and examining the past few years of evaluation results, Trevor gathers his team for strategic planning.

After several weeks of researching options, Trevor’s team decides to improve their evaluations by:

• Reframing all their survey questions to be learner-centric (stated from learner’s viewpoint).
• Focusing particularly on how likely learners are to recommend the learning to others.

How would you rate the effectiveness of Trevor’s team’s new measurement approach as an indicator of learning success?

A. NOT AT ALL effective.
B. SOMEWHAT effective.
C. MOSTLY effective.
D. VERY effective.

Sharon Shrock and Bill Coscarelli, authors of the classic text, now in its third edition, *Criterion-Referenced Test Development*, offer the following wisdom:

On using Likert-type Descriptive Scales (of the kind that uses response words such as “Agree,” “Strongly Agree,” etc.):

“...the resulting scale is deficient in that the [response words] are open to many interpretations.” (p. 188)
Research shows that learners don’t always know their own learning...

**Learners are Overly Optimistic**
Zechmeister & Shaughnessy (1980).

**Learners Fail to Properly Use Examples**
Renkl (1997).

**Learners can’t always Overcome Faulty Prior Knowledge**
Kendeou & van den Broek (2005).

**Learners Fail to Give Themselves Retrieval Practice**
Karpicke, Butler, & Roediger (2009).

**Two Recent Reviews Emphasize Learners’ Lack of Knowledge of Learning**

Trevor was recently promoted into a new position, the Director of Learning Evaluation for an integrated health network.

**Measured:**

- Learner Satisfaction
- Reputation of the Learning
HOW ABLE ARE YOU to put what you’ve learned into practice on the job? Choose One.

A. I am NOT AT ALL ready to use the skills taught.
B. I have GENERAL AWARENESS but will NEED MORE GUIDANCE to put the skills into practice.
C. I need MORE HANDS-ON EXPERIENCE to be GOOD at using these skills.
D. I am FULLY COMPETENT in using these skills.
E. I am CAPABLE at an EXPERT LEVEL in using these skills.
After the course, when you begin to apply your new knowledge at your worksite, which of the following supports are likely to be in place for you?
Select as many items as are likely to be true.

A. I will have my PROGRESS MONITORED BY MY SUPERVISOR in applying the learning.
B. I will have someone available TO COACH OR MENTOR ME in applying the learning.
C. I will have easy access to a COURSE INSTRUCTOR to contact for guidance and support.
D. I will have JOB AIDS to guide me in applying the learning to real job tasks.
E. I will be PERIODICALLY REMINDED of key learning concepts/skills over the next few months.
F. I will NOT get much direct support, but will rely on my own initiative.

Percent saying NEW QUESTIONS BETTER than traditional questions:

80%
<table>
<thead>
<tr>
<th>Three Open-Ended Questions to End Your Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>What aspects of the training made it MOST EFFECTIVE FOR YOU? What should WE DEFINITELY KEEP as part of the training?</td>
</tr>
<tr>
<td>What aspects of the training COULD BE IMPROVED? Remember, your feedback is critical, especially in providing us with constructive ideas for improvement.</td>
</tr>
<tr>
<td>Is there anything else we should have asked about? Is there anything you want to tell us?</td>
</tr>
</tbody>
</table>

I am clear about what is expected of me as a result of going through this training.

A. Strongly Disagree
B. Disagree
C. Neutral
D. Agree
E. Strongly Agree
How READY are you TO WRITE SCENARIO-BASED QUESTIONS?

CIRCLE ONE OR MORE ANSWERS AND WRITE YOUR REASONING BELOW

A. I’m STILL NOT SURE WHERE TO BEGIN.
B. I KNOW ENOUGH TO GET STARTED.
C. I CAN TELL A GOOD QUESTION FROM A BAD ONE.
D. I CAN WRITE MY OWN QUESTIONS, but I’d LIKE to get SOME FEEDBACK before using them.
E. I CAN WRITE MY OWN QUESTIONS, and I’m CONFIDENT they will be reasonably WELL DESIGNED.

In your own words, how ready do you feel you are in being able to write scenario-based questions?

What’s Wrong?

After training, my manager and I will discuss how I will use the learning on my job.

strongly disagree  1 2 3 4 5 6 7 strongly agree
What plans, if any, do you have for talking with your manager in the next 10 days about how you will use the learning in your work?

A. My manager and I have made plans for at least TWO meetings.
B. My manager and I have made plans for at least ONE meeting.
C. My manager HAS MENTIONED the idea, but we don’t yet have firm plans.
D. I will SEEK OUT my manager and ask for at least one meeting on this.
E. We are LIKELY TO DISCUSS my use of the learning as we work together.
F. It is DOUBTFUL that we will spend much time discussing my use of the learning.

Better!

Smile Sheets should be ONLY ONE PART of our learning evaluation efforts

- Smile Sheets
  - Understanding
  - Remembering
  - Motivation to Apply
  - After Supports

- Learners able to:
  - Understand?
  - Remember?
  - Make Decisions?
  - Apply the Learning?

- Supports:
  - Management Support?
  - Workplace Obstacles?
  - Reinforcement?
  - Reminders?

- Meeting Target Goals?
  - Job Performance
  - Organizational Results
  - Learner Expectations
  - Other Expectations
Performance-Focused Smile Sheets:
What Questions Do You Have?

Recommended Process for Writing Questions

1. Gather example performance-focused smile sheet questions.
2. Determine what messages you want to send with your questions.
3. Target at least one question focused on:
   - Learner comprehension
   - Learner motivation to apply
   - Support for remembering
   - Support for after-learning follow-up
4. Engage in a cycle of question writing, expert feedback and question writing, and rewriting until questions are ready.
5. Get stakeholder agreement to pilot test.
6. Pilot test on one (or a few) learning interventions.
7. Review results, engage with stakeholders who will make changes to learning, get expert feedback, improve questions.
8. Deploy questions, periodically reviewing them, or replacing some to focus your team on a particular set of learning factors.
## Tier 5

### Learner makes decisions given relevant realistic scenarios.
- **Decision Making Competence**—during or right after learning event.
- **Remembered Decision Making Competence**—after several days or more.

### The Learning-Transfer Evaluation Model

<table>
<thead>
<tr>
<th>Tier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Effects of Transfer</strong></td>
</tr>
<tr>
<td>2</td>
<td><strong>Transfer</strong></td>
</tr>
<tr>
<td>3</td>
<td><strong>Task Competence</strong></td>
</tr>
<tr>
<td>4</td>
<td><strong>Decision Making Competence</strong></td>
</tr>
<tr>
<td>5</td>
<td><strong>Knowledge</strong></td>
</tr>
<tr>
<td>6</td>
<td><strong>Learner Perceptions</strong></td>
</tr>
<tr>
<td>7</td>
<td><strong>Activity</strong></td>
</tr>
<tr>
<td>8</td>
<td><strong>Attendee</strong></td>
</tr>
</tbody>
</table>

**Tier 5**

**Decision Making Competence**

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Learner makes decisions given relevant realistic scenarios.</td>
</tr>
<tr>
<td>2</td>
<td><strong>Decision Making Competence</strong>—during or right after learning event.</td>
</tr>
<tr>
<td>3</td>
<td><strong>Remembered Decision Making Competence</strong>—after several days or more.</td>
</tr>
</tbody>
</table>

**ADVERSE TO CERTIFY DECISION MAKING COMPETENCE.**
Some Terminology

Questions can sometimes do both…

Questions For Testing

- May Require Same Number of Answer Choices
- May Need Rigorous Validation
- May Need an Item Pool
- May Need Test-Taking Security

Questions For Learning

- More Flexibility
- Can Incorporate Comprehensive Feedback
- Can be More Fun and Engaging
- Support Long-Term Remembering
Your employees attend a four-day course on video cinematography.

The following week, they will be given a paper-and-pencil certifying exam, which will be proctored in your organization’s highly-distinctive Oak Boardroom.

If you want to maximize your learners’ passing rate, where should you hold your four-day course?

A. In the Oak Boardroom.
B. In any distraction-free room.
C. In a room nearby to their workspace.
D. In a room with video/lighting equipment and computers set up to edit video/audio.
How Learners Encode Learning Stimuli

Background Context

Fosse
Ditch

How Background Stimuli Triggers Retrieval of Learned Information

Cue
Ditch
Fosse

Action

Long Term Memory

Background Context


Some Research on Context Alignment


Retrieval = Learning – Forgetting + Spontaneous Remembering
How to Create:

**Spontaneous Remembering**

#1 Thing to Do
Change learning context to make it similar to performance context

Hands-on  Simulation  Scenarios

From Situation to Action

Situation  Fosse  Ditch  Action

Background Context

Long Term Memory
From Situation to Action

Situation

Fosse

Ditch

Background Context

Long-Term Memory

Action

SEDA Model

Situation

Evaluation

Decision

Action
Example: Teaching Management

**Topic-Based**
- You need to bring your direct reports into decision-making to increase their buy-in.

**Situation-Based**
- Bring your direct reports into decision making to increase their buy-in:
  - In staff meetings
  - In project planning
  - When brainstorming ideas
- But not:
  - In urgent situations
  - For safety, legal issues

Training-Design Breakdowns

*Brought to light with the SEDA Model*

- **No Situation**
  - Teaching Concepts w/o Practice

- **Inadequate Evaluation**
  - Providing Hints through Chapterization.

- **No Decision, No Action**
  - Case Studies w/o real decisions

**Role Play Example:**
"Customer is angry because of poor product performance."
"Express empathy and follow four step empathy process."
Simulation-Design Breakdowns
Brought to light with the SEDA Model

**Situation**

- **Non-Realistic**
  - Set in Non-Realistic Situations: "King Arthur"

- **Stimuli Overload**
  - Distractions too early in learning.

- **Too Soon After Learning**
  - Not Realistic Evaluation/Decision

**Evaluation**

- **Bounded Content Domain**
  - Providing Chapterization Hints: "Heart valve-problem simulation"

- **Closed Problem Space**
  - Multiple-Choice Answers

- **No Evaluation**
  - Simulation w/Hints: "Patient is alcoholic..."

**Decision**

- **No Practice in Taking Action**
  - Where simulation focus is on decision-making, not doing the action per se.

**Action**

SEDAModel
First Step In Using the Model in Learning Design

*What are the situations in which we want our learners to remember what they learned from us?*
What do learners need to be able to do, and in what situations do they need to do those things?

The Magic Question:

“Your Missed Opportunities?”

<table>
<thead>
<tr>
<th>SEDA Issue</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 No Situation</td>
<td>Provide no/few situations for learners to respond to, practice with.</td>
<td>Just presenting information.</td>
</tr>
<tr>
<td>2 No Situation</td>
<td>Provide tests of memorization instead of authentic decision making.</td>
<td>Asking about definitions, terminology, or using Bloom's.</td>
</tr>
<tr>
<td>3 No/Poor Evaluation</td>
<td>Give them situation, but tell them what it means, what category it is, etc.</td>
<td>“The patient is an alcoholic...”</td>
</tr>
<tr>
<td>4 No/Poor Evaluation</td>
<td>Chapterize concepts and only give info/practice within those chapters (not across chapters).</td>
<td>Testing on quizzes or knowledge checks w/ no cumulative exam.</td>
</tr>
<tr>
<td>5 Situation/ Evaluation Only</td>
<td>Asking learners only to evaluate situations without having them make authentic decisions and/or take actions.</td>
<td>Case studies that don't ask people to make decisions about what to do.</td>
</tr>
<tr>
<td>6 Too-Generic Situation</td>
<td>Asking learners to deal with a situation, but one that is too generic to have effective cues.</td>
<td>“What should Joe do at his next staff meeting?”</td>
</tr>
<tr>
<td>7 No Situation No Evaluation</td>
<td>Showing examples, without linking those examples to when they might be needed.</td>
<td>Showing people how to hide a column in Excel, w/o describing when this would be valuable.</td>
</tr>
<tr>
<td>8 No Evaluation No Decision</td>
<td>Just giving people practice on actions without requiring them to make critical evaluations or decisions.</td>
<td>“With your role-play partner, practice your active listening skills after giving feedback regarding the person's lack of effort.”</td>
</tr>
</tbody>
</table>
Let’s Analyze Question Quality

You work as a consultant/salesperson making recommendations to business customers.

Verdeblock is especially good for copper oxidation, but it should NOT be used on surfaces from which people may eat.

Grayblock is especially good for tarnished silver, but it can be used for other oxidation. It is safe for eating surfaces.

Verdeblock is particularly beneficial in reducing which of the following:

A. Copper oxidation.
B. Lime deposits.
C. Silver tarnish.
D. All of the above.

INFORMATION:

Verdeblock is especially good for copper oxidation, but it should NOT be used on surfaces from which people may eat.

Grayblock is especially good for tarnished silver, but it can be used for other oxidation. It is safe for eating surfaces.

SEDA Analysis
You’re with a customer who makes furniture that people eat off of. The furniture has substantial copper, which is oxidizing. Which product do you recommend for oxidation on eating surfaces?

A. Suggest Verdeblock.
B. Suggest Grayblock.
C. Suggest Blublock
D. Any of the above are okay.

**INFORMATION:**

Verdeblock is especially good for copper oxidation, but it should NOT be used on surfaces from which people may eat.

Grayblock is especially good for tarnished silver, but it can be used for other oxidation. It is safe for eating surfaces.

You’re with a customer who makes patio furniture. You observe her inventory and see that the furniture has substantial copper, which is oxidizing. Which product do you recommend?

A. Suggest Verdeblock.
B. Suggest Grayblock.
C. Gather more information.
D. None of the above.

**INFORMATION:**

Verdeblock is especially good for copper oxidation, but it should NOT be used on surfaces from which people may eat.

Grayblock is especially good for tarnished silver, but it can be used for other oxidation. It is safe for eating surfaces.
You're with a customer who needs Graylock because she has an oxidation problem on a copper eating surface. Show how you would recommend Graylock to her by speaking into the microphone. Say the following:

“Our best product for copper oxidation on food surfaces is Grayblock. It works great on copper and it’s absolutely safe. How many cases will you need?”

INFORMATION:

Verdeblock is especially good for copper oxidation, but it should NOT be used on surfaces from which people may eat.

Grayblock is especially good for tarnished silver, but it can be used for other oxidation. It is safe for eating surfaces.
### Task Competence

Learner performs relevant realistic actions and decision making.
- **Task Competence**—during or right after learning event.
- **Remembered Task Competence**—after several days or more.

*ADEQUATE TO CERTIFY TASK COMPETENCE.*

**NOTE:** "Tasks" comprise both decision making and action taking. For example, a person learning to write poetry could decide to use metaphor, could get to use it, or could do both.

### Decision Making Competence

Learner makes decisions given relevant realistic scenarios.
- **Decision Making Competence**—during or right after learning event.
- **Remembered Decision Making Competence**—after several days or more.

*ADEQUATE TO CERTIFY DECISION MAKING COMPETENCE.*
From a Learning Standpoint, What are the downsides to Multiple-Choice Questions?

1. The Problem Space is Confined
2. The Choices are Not Fully Processed
One-Option-at-a-Time Approaches

• Standard Questions
  Supporting "Evaluation" of "Situation" by using complex situation, where only someone who knows what to look for will do well.

• None-of-the-Above Questions
  If None-of-the-Above is chosen, branch to other decision.

• Two-Part, Evaluation-First Design
  First scene asks if a decision made in the scenario was a good one, or first scene asks if you have enough information to make a decision. After first scene, branch to Decision after giving within-the-fiction feedback on the evaluation chosen.

• Two-Part, Forced-Evaluation Design
  First scene asks learner to specifically evaluate the situation. Branch to Decision after giving within-the-fiction feedback on the evaluation of the situation.

• Multi-Scene Designs (For example: Data-gathering, Diagnosis, Decision)
  For example, first scene might ask which additional data to gather (to better understand the Situation), second scene might ask for a diagnosis on the problem, third scene might ask for a selection of a solution.

---

Example of:

One Option at a Time Scenario

Alena wants to start a firm that helps farmers grow food organically. She has a degree in sustainable agriculture and has worked for a non-profit organization for seven years doing similar work. She’s developed a marketing plan, a financial plan, and has found several farmers who would pay her if she went out on her own. What should Alena do first—before she tells her boss that she’s quitting to start her own firm?

A. Create a cash flow statement to determine whether her predicted income will support the business through the first year.
B. Form a group of advisors with experience in both small-business management and agriculture.
C. Determine whether she has enough seed money and start-up capital to get started.
D. Analyze her values and goals to ensure that the proposed business will support them.
Simultaneous Multiple-Option Approaches
ProvidingMenus, Selectable Objects, Immersive Environments that enable:

1. Situation Data-Gathering
2. Evaluation-Testing
3. Decision-Making
4. Action Taking (or simulated action-taking)

Example of:
Multiple Options at a Time Scenario/Simulation
### Simultaneous Multiple-Option Approaches

Providing Menus, Selectable Objects, Immersive Environments that enable:

1. Situation Data-Gathering
2. Evaluation-Testing
3. Decision-Making
4. Action Taking (or simulated action-taking)

**Major Advantage of Multiple-Option Approaches**

- MORE REALISTIC because learners have to decide how to engage.
  - They have to determine whether more situation-information is needed,
  - whether evaluations have to be tested,
  - what decisions should be made, etc.

**Major Disadvantages of Multiple-Option Approaches**

- May overload learner’s working-memory capacity.
- May cause frustration.
- Much more difficult to design.
- FOR EVALUATION: Extremely difficult to validate.

### One-Option-at-a-Time Approaches

- **Standard Questions**
  Supporting “Evaluation” of “Situation” by using complex situation, where only someone who knows what to look for will do well.

- **None-of-the-Above Questions**
  If None-of-the-Above is chosen, branch to other decision.

- **Two-Part, Evaluation-First Design**
  First scene asks if a decision made in the scenario was a good one, or first scene asks if you have enough information to make a decision. After first scene, branch to Decision after giving within-the-fiction feedback on the evaluation chosen.

- **Two-Part, Forced-Evaluation Design**
  First scene asks learner to specifically evaluate the situation. Branch to Decision after giving within-the-fiction feedback on the evaluation of the situation.

- **Multi-Scene Designs (For example: Data-gathering, Diagnosis, Decision)**
  For example, first scene might ask which additional data to gather (to better understand the Situation), second scene might ask for a diagnosis on the problem, third scene might ask for a selection of a solution.
You were recently hired as a store manager for a retail drug store that just opened. Your pharmacy staff are experienced, but your retail staff is mixed.

You have experienced store clerks but two relatively inexperienced assistant managers. Dave has worked as a sales clerk in the company for 5 years, and just became an assistant store manager 2 months ago. Deedee has 6 years of clerking experience, and 6 months as an assistant manager.

Your store has been opened for a month when your district manager swings by. She walks the store with you and points out more merchandising mistakes than is acceptable. You feel ashamed and appalled because you can see clearly what is wrong.

What is your best course of action to correct the problem?

A. Walk the store yourself, have your assistant managers observe you as you talk them through the process and develop the task list.

B. Let them take turns walking the store and developing the task list. Watch them closely and give them feedback as they write the task list.

C. Let them take turns walking the store and developing the task list. Check the task list when they are done and then give them feedback.

D. Show them how to walk the store. Soon, have them walk the store while you give feedback. Later, have them do it solo; coach as needed.

One-Option-at-a-Time Approaches

- **Standard Questions**
  Supporting “Evaluation” of “Situation” by using complex situation, where only someone who knows what to look for will do well.

- **None-of-the-Above Questions**
  If None-of-the-Above is chosen, branch to other decision.

- **Two-Part, Evaluation-First Design**
  First scene asks if a decision made in the scenario was a good one, or first scene asks if you have enough information to make a decision. After first scene, branch to Decision after giving within-the-fiction feedback on the evaluation chosen.

- **Two-Part, Forced-Evaluation Design**
  First scene asks learner to specifically evaluate the situation. Branch to Decision after giving within-the-fiction feedback on the evaluation of the situation.

- **Multi-Scene Designs (For example: Data-gathering, Diagnosis, Decision)**
  For example, first scene might ask which additional data to gather (to better understand the Situation), second scene might ask for a diagnosis on the problem, third scene might ask for a selection of a solution.
None-of-the-Above Approach
Question 1

- You were recently hired as a store manager for a retail drug store that just opened. Your pharmacy staff are experienced, but your retail staff is mixed.
- You have experienced store clerks but two relatively inexperienced assistant managers. Dave has worked as a sales clerk in the company for 5 years, and just became an assistant store manager 2 months ago. Deedee has 6 years of clerking experience, and 6 months as an assistant manager.
- Your store has been opened for a month when your district manager swings by. She walks the store with you and points out more merchandising mistakes than is acceptable. You feel ashamed and appalled because you can see clearly what is wrong.
- What is your best course of action to correct the problem?

A. Walk the store yourself, have your assistant managers observe you as you talk them through the process and develop the task list.
B. Let them take turns walking the store and developing the task list. Watch them closely and give them feedback as they write the task list.
C. Let them take turns walking the store and developing the task list. Check the task list when they are done and then give them feedback.
D. There is a significantly better approach than any of the answers above.

None of the Above
Two-Question Sequence

Scenario

Learning Point
Question 2

- After you made your decision, your boss comes in and tells you to reflect on the issue in a broader more-expansive way.
- Given this helpful nudge, you consider some options. What action will you take with your assistant managers to improve your store’s appearance?

A. Show them how to walk the store. Give them a list of things to focus on as they walk the store. Follow-up several times over the next few days.

B. Show them how to walk the store. Soon, have them walk the store while you give feedback. Later, have them do it solo; coach as needed.

C. Have them walk the store. Look to see what they do right and wrong. Provide them with corrective feedback. Monitor until they improve.

One-Option-at-a-Time Approaches

- Standard Questions
  Supporting “Evaluation” of “Situation” by using complex situation, where only someone who knows what to look for will do well.

- None-of-the-Above Questions
  If None-of-the-Above is chosen, branch to other decision.

- Two-Part, Evaluation-First Design
  First scene asks if a decision made in the scenario was a good one, or first scene asks if you have enough information to make a decision. After first scene, branch to Decision after giving within-the-fiction feedback on the evaluation chosen.

- Two-Part, Forced-Evaluation Design
  First scene asks learner to specifically evaluate the situation. Branch to Decision after giving within-the-fiction feedback on the evaluation of the situation.

- Multi-Scene Designs (For example: Data-gathering, Diagnosis, Decision)
  For example, first scene might ask which additional data to gather (to better understand the Situation), second scene might ask for a diagnosis on the problem, third scene might ask for a selection of a solution.
Elsa was recently hired as a store manager for a retail drug store that just opened. Her pharmacy staff are experienced, but her retail staff is mixed.

She has experienced store clerks but two relatively inexperienced assistant managers. Dave has worked as a sales clerk in the company for 5 years, and just became an assistant store manager 2 months ago. Deedee has 6 years of clerking experience, and 6 months as an assistant manager.

Elsa’s store has been opened for a month when her district manager swings by. Ms. Hawkeye walks the store with Elsa and points out more merchandising mistakes than is acceptable. Elsa appears ashamed and appalled and tells Ms. Hawkeye that she’ll get the problem fixed right away.

Elsa decides to walk the store twice a day, taking her assistant managers with her, showing them what is good merchandising and what is poor. Her plan is to do this for a week or two—or until the problem is fixed.

If you were hired to coach Elsa, what would you recommend?

A. Coach Elsa that her plan is a good starting point.
B. Coach Elsa that her plan needs improvement.
Before Elsa has time to act on your recommendation, Ms. Hawkeye, District Manager Extraordinaire, learns of Elsa’s plan—and she does NOT like it. She coaches Elsa that her plan is inadequate and gives her a few pointers before telling Elsa to speak with you further.

Knowing that you are now on the hot seat—and your lucrative consulting gig is in severe jeopardy—what recommendation do you make to Elsa?

Do you tell Elsa to:

A. Walk the store yourself, have your assistant managers observe you as you talk them through the process and develop the task list.

B. Let them take turns walking the store and developing the task list. Watch them closely and give them feedback as they write the task list.

C. Let them take turns walking the store and developing the task list. Check the task list when they are done and then give them feedback.

D. Show them how to walk the store. Soon, have them walk the store while you give feedback. Later, have them do it solo; coach as needed.

What Questions do You Have?
How to Write Scenario-Based Questions (For Evaluation)

1. Learn deeply about the performance needs and situations, the stakeholders, the content, and the articulated learning goals.
2. Outline a process for question design and development. Get buy-in on it.
3. Answer the magic question: "What do learners need to be able to do, and in what situations do they need to do those things?"
4. Specify a clear set of learning goals.
5. Get buy-in on the learning goals from key stakeholders.
6. Write first draft of questions.
7. Get feedback from one or more experts, then rewrite or improve.
8. Validate your questions by having experts answer your questions THEN comment on them. (See Validation)
9. Improve questions, validate again depending on criticality.
10. Clone questions (write versions with different salient information, including different background information).
11. Randomly assign originals and clones to different assessments (pretests, posttests, delayed tests, item pools).
12. Pilot and improve as needed.

Deep Analysis to Create Learning Goals
Iterations Toward Improved Questions
Validation of Your Questions
Clone Questions
Pilot

Gertrude has spent two years learning about and getting to know many of the nine doctors at Lexington Dermatological Associates. She knows most of them fairly well and enjoys meeting with them, though she wishes she had more time to spend so she could increase sales of her company’s products. Recently, after attending a brilliant training program (aren’t they all?), Gertrude redoubled her efforts to build stronger relationships. In her next meeting with Dr. Speckles, the director, she focuses on creating stronger bonds while also emphasizing key sales messages that her sales team talked about in their most recent meeting.

Suppose you were Gertrude coach, what would you tell her about her performance? What’s the most important feedback you can provide?

A. Gertrude, you have the basics down, but you need to ask more questions and focus more on helping your customers help their patients.
B. Gertrude, a good start, however, you need to be more interactive and ask more questions rather than just emphasizing your selling messages.
C. Gertrude, good work! Your emphasis on relationship-building will build trust and help you make your sales targets.
D. Gertrude, great work! You went the extra mile in building tighter bonds with a key customer.

Validating Your Questions
(An Example)

Best Answer
Second Best

Learning Point:
A top priority is to help our customers help their patients by providing them with what they really need. We need to ask diagnostic questions so we can help.
Gertrude has spent two years learning about and bonding with the nine doctors at Lexington Dermatological Associates. She knows each person fairly well and enjoys meeting with them, though she wishes she had more time to spend so she could increase sales of her company’s products. Recently, after attending a brilliant training program, Gertrude redoubled her efforts to build stronger relationships. In her next meeting with Dr. Speckles, she focuses on creating stronger bonds while also emphasizing key sales messages that her sales team talked about in their most recent meeting.

Suppose you were Gertrude coach, what would you tell her about her performance? What’s the most important feedback you can provide?

A. Gertrude, you have the basics down, but you need to ask more questions and focus more on helping your customers help their patients.
B. Gertrude, a good start, however, you need to be more interactive and ask more questions rather than just emphasizing your selling messages.
C. Gertrude, good work! Your emphasis on relationship-building will build trust and help you make your sales targets.
D. Gertrude, great work! You went the extra mile and were able to build tighter bonds with a key customer.

How to Validate:

1. Deeply understand the need and outline clear learning points.
2. Write questions, get input from experts to improve.
3. Find a handful of new experts and ask them to answer your questions without providing them with any information or hints.
4. After that, ask them to explain their rationale for their answers.
5. Improve questions & answer choices.
6. Test again depending on the criticality of the assessment. Consider testing with target population.
7. Pilot and improve.

Practice with Feedback from Will

1. Write one or both of the following within 10 days from now:
   - Learner-Feedback Question
   - Scenario-Based Question
2. Send to me at: will.thalheimer@worklearning.com
3. I’ll send you feedback within 10 days.
Overarching Messages

1. We should do evaluation to help us make our most important decisions.

2. As an industry, we are not satisfied with our learning-evaluation practices. Clearly then, some or many things that we are doing are insufficient or problematic.

3. Many of our problems have arisen because we haven’t taken into account human learning and cognition.

4. Our traditional learner-feedback methods are not adequate. The performance-focused smile sheet methodology can help.

5. Our traditional evaluation models are not adequate. Utilizing LTEM can help.

6. Learning evaluation is very complex and difficult, so we should be realistic and humble in our efforts.

7. By doing better learning evaluation, we will be able to improve our learning interventions and give our organizations a competitive advantage.

Specific Tools and Insights (Page 1)

1. Data should be accurate, valid, relevant, highly predictive, cost effective, and help us make our most important decisions.

2. Measurement illuminates some things, but leaves other things uninvestigated and ignored.

3. There is no perfect measurement approach, system, tool, or practice.

4. We are looking for reasonable proxies for the constructs we are hoping to measure.

5. What is most important in learning measurement is looking for ways to maintain what is good and improve what is bad in our learning interventions and practices. Sadly, this has often been overlooked as a goal for us.

6. Remembering (and minimizing forgetting) is a crucial goal for learning design. Therefore, we should evaluate remembering.

7. Measuring right at the end of learning will tend to produce biased results—results that will appear better than what our actual learning results will be.

8. Measuring in the learning context will likely produce biased results—results that will look better than our actual learning results.

9. Measuring low-level knowledge of concepts will generally produce biased results—results that don’t capture the capability of our learners to perform skills, make decisions, and complete real-world tasks.
10. Raymond Katzell was the originator of the Four-Level Model, while Donald Kirkpatrick was the force behind its popularization.

11. The Kirkpatrick-Katzell Four-Level Model sends both helpful messages and harmful messages, and has contributed to many of our inadequate evaluation practices.

12. LTEM was specifically designed to send messages that are helpful—to nudge us toward better evaluation practices.

13. LTEM incorporates wisdom from fundamental aspects of human learning and cognition.

14. LTEM is not a panacea.

15. LTEM can be used in two ways: (A) As a map of what is possible that enables us to look to see what we are doing now and what improvements we can make, (B) As a guide in our development work, enabling us to work backward from our goals to our evaluation practices.

16. All of our models and practices send messages and nudge action, whether intentionally sent or not; whether consciously or subconsciously received.

17. Traditional smile sheets are virtually uncorrelated with learning results and are thus unable to guide us in making learning-design improvements.

18. Performance-focused smile-sheet questions are better because they support learner decision-making, are more motivating to respondents, and produce data that has greater clarity and is more actionable.

19. Likert-like scales and numeric scales are too fuzzy to enable good decision-making and actionable data.

20. Learners are not always accurate in thinking about what works best in learning, so any calls to be learner-focused should be interpreted with caution.

21. Smile sheets should NOT be the only method we use in evaluating learning.

22. Open-ended smile sheet questions can provide rich information, especially when they are preceded by performance-focused smile sheet questions that encourage a focus on important learning factors.

23. Questions that support learning and questions that enable evaluation have differing requirements.

24. Human beings are greatly influenced by cues in situational contexts. Indeed, thoughts and actions often flow directly from situations.

25. When we utilize the same or similar cues during learning that learners will face in their work situations, we can help trigger remembering and action when learners later encounter those cues in their work.
26. One of the most important things we can do to support remembering is to utilize the context alignment notion, changing the learning context to make it similar to the learners’ future work contexts.

27. Our goal as learning designers should be to increase learning, minimize forgetting, and enable spontaneous remembering.

28. The SEDA model nicely captures the learning science notion of context alignment.

29. SEDA represents the idea that people do their work in SITUATIONS, make constant EVALUATIONS of those situations, make DECISIONS based on their evaluations, and take ACTIONS to implement those decisions.

30. By using SEDA as a guide in analyzing our learning designs, we can see what is missing in our designs.

31. Too often, we fail to provide our learners with a full SEDA experience.

32. The Magic Question helps us to focus on the situational nature of performance.

33. The SEDA Model is helpful in writing scenario-based questions.

34. Tier 5 can be assured with a SED approach, whereas Tier 6 requires a full SEDA approach.

35. Multiple choice questions too often close the problem space, short-circuiting the cognitive processing that learners have to do—and thus harming learning.

36. Scenarios can prompt learners to decide one thing at a time or choose between multiple decision options.

37. Multiple-choice questions can close the problem space, but there are a number of methods that can ameliorate these issues to some extent.

38. Writing scenario-based questions for evaluation may require as many as 12 steps.

39. Scenario-based questions used for evaluation should ideally be validated.

40. One approach to validation involves having experts answer the questions and then discussing their thought process in thinking through the scenario.
The Learning-Transfer Evaluation Model (LTEM) and Fundamental Secrets in Learning Evaluation

Will Thalheimer, PhD

Phone: +1-617-718-0767
Email: info@WorkLearning.com
Website: WorkLearning.com
Book: SmileSheets.com
Twitter: @WillWorkLearn

Slides available at:
www.is.gd/will999stuff
Three Reasons to Use Measurement in Learning

1. To provide learners’ managers with feedback and information.
2. To provide other organizational stakeholders with information.
3. To examine the organizational impacts of learning.
4. To compare one learning intervention to an alternative one.
5. To calculate ROI of the learning program.
6. To collect data to sell or market the learning program.
7. To assign learners with grades or give them a passing score.
8. To enable learners to earn credentials.
9. To document legal or regulatory compliance.
Three Reasons to Use Measurement in Learning

1. To encourage learners to study.
2. To diagnose learning needs.
3. To give learners feedback on their learning progress.
4. To help learners better understand the concepts being taught by giving them tests of understanding and follow-up feedback.
5. To provide learners with additional retrieval practice (to support long-term retrieval).
6. To give successful assessment-takers a sense of accomplishment, a sense of being special, and/or a feeling of being in a privileged group.
7. To increase the likelihood that the learning is implemented.

Three Reasons to Use Measurement in Learning

1. To provide instructors with feedback on learning.
2. To provide instructional designers/developers with feedback.
3. To diagnose future learning needs.
4. To compare the effectiveness of one learning method to another learning method.
5. To diagnose workplace obstacles to full application.
Tier 6 Best Practices

1. Before designing learning and before determining instructional objectives, first determine how you will evaluate at each Tier.

2. For Tier 6, specify these “evaluation objectives” to measure how well your learners can do the tasks (or subset of tasks) that you hope they can do later in their work. In other words, design your Tier 6 exercises before designing your learning!

3. Develop methods that will help observers (or technology-based monitors) evaluate fairly, minimizing bias—for example easy guidelines, simple rubrics, clear criterion for success. Pilot test these to improve them.

4. Where possible, focus your attention on evaluating the learning design more than the learners. To your learners, frame it as a learning experience rather than a high-stakes assessment.

5. Make the task performance more authentic by “testing” after a delay, by minimizing artificial hints and supports, and by avoiding exact repetitions of previously encountered task stimuli. Note that if people “study” for your assessment by cramming, your results will not be fully authentic.
Tier 7 Data-Gathering Methods

- Surveys
- Interviews
- Focus Groups
- Observations
- End-Product Reviews
- Prepared Observations
- Behavioral Metrics
- Validated Behavioral Metrics

Estimated Best Guess

LOWER
Validity/Reliability

MEDIUM

HIGHER
Validity/Reliability

Tier 7 Best Practices

1. Use Comparison Group, if possible
2. Compare Pre and Post, if possible
3. Utilize Multiple Data Points...
4. Measure at Multiple Time Intervals...
5. Use Data that is More Reliable/Valid
6. Build Partnership with Organizational Stakeholders
7. Consider both Costs and Benefits
8. Examine Supports/Obstacles for Work Application
9. Be Practical and Evaluate at Reasonable Costs
1. Complete Tier 7 Evaluation, then consider similar best practices for Tier 8, including the use of comparison groups, pre and post testing, multiple data points and measuring over time.

2. Use data that is more valid/reliable and that can show transfer’s causal impact.

3. Consider focusing not just on results impacting the organization, but on all of these: learners, coworkers/family/friends, the organization, community, society, and the environs.

4. Look at both positive and negative effects of the learning intervention.