Enlightening Ourselves as Learning Professionals

Will Thalheimer, PhD

Research Translation

WorkLearning.com

Bridging Gap between Research and Practice
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?

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

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

What are our most important decisions?

What about our smile-sheet (learner-feedback) data?

What about our quiz and knowledge-check data?

What about our post-training on-the-job performance data?
Weak Relationship is below .30 and .09 is VERY WEAK

So...SMILE SHEETS tell us VERY LITTLE about Learning


1990’s

2000’s

Remembering
The Learning Landscape

Learning Intervention

Performance Situation

Learning Outcomes

Learning

Remembering

On-the-Job Performance

Individual Results

Organizational Results

Learning and Forgetting Curves

More Remembering

Less Remembering

During Learning

After Learning

© Copyright by Work-Learning Research, Inc. (www.work-learning.com)
Learning and Forgetting Curves

What does an end-of-course assessment tell us?

YouTube: http://is.gd/LearningForgettingCurves
Learning and Forgetting Curves

Understanding

During Learning

After Learning

Delayed Tests

Remembering

Application

Biased Metric!!

If we measure here?

Understanding

Remembering

Application

Delayed Tests

Biased Metric!!

Understanding

When did you measure learning in question 4, when were the evaluation instruments presented to the learners? Check all that apply.

- At the end of the learning event
- Later the same day
- On the next day after the learning event ended
- From 2 to 6 days later
- From 1 to 3 weeks later
- From 1 to 3 months later
- From 4 months to 1 year later
- More than 1 year later

Source: The Learning Guild Research
WHERE did you Measure Learning?

7) If you said you measured Learning in Question 4, HOW SIMILAR was the ASSESSMENT CONTEXT to the LEARNING CONTEXT?

Mostly in the Same Context: 53.9%
Mostly in Similar Contexts: 37.1%
Mostly in Different Contexts: 6.6%
Not Applicable: 3.0%

Source: The Learning Guild Research.

Tested in the learning room or in a different room. Does this matter?

"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

<table>
<thead>
<tr>
<th>Percent of Companies Using “To Any Extent”</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATD and i4cp</td>
</tr>
<tr>
<td>Classroom – Reaction 52%</td>
</tr>
<tr>
<td>Classroom – Learning 50%</td>
</tr>
<tr>
<td>Technology – Reaction 52%</td>
</tr>
<tr>
<td>Technology – Learning 43%</td>
</tr>
<tr>
<td>Behavior 15%</td>
</tr>
<tr>
<td>Results 10%</td>
</tr>
<tr>
<td>ROI 6%</td>
</tr>
</tbody>
</table>

ATD and i4cp
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

97% want to improve the way they gather and analyse data on learning impact

“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-Katzell Four-Level Model

Level 1: Reaction
Level 2: Learning
Level 3: Behavior
Level 4: Results

Measurement illuminates some things.

Measurement doesn’t illuminate everything.

Read about Raymond Katzell’s role:
https://is.gd/originator
"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!

"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!!

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.

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.
Likert-like Scales provide Poor Data

A. Strongly Agree
B. Agree
C. Neither Agree Nor Disagree
D. Disagree
E. Strongly Disagree

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

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)
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
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

A Better Smile Sheet Question

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.
HOW ABLE ARE YOU to put what you’ve learned into practice on the job? Choose One.

- NOT AT ALL READY: Deplorable
- HAVE AWARENESS, NEED MORE GUIDANCE: Unacceptable
- NEED MORE HANDS ON EXPERIENCE: Acceptable
- FULLY COMPETENT: Superior
- COMPETENT AT EXPERT LEVEL: Unlikely

Percentage of Respondents

A Question About Follow-Through...

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.
But won’t learners dislike these new type of questions?

Percent saying NEW QUESTIONS BETTER than traditional questions:

80%
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
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>Level 2 Learning is Mashed into One Bucket</td>
<td></td>
</tr>
</tbody>
</table>

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

- Just because learners ENGAGE IN LEARNING doesn’t mean they will have learned. Therefore, measuring attendance is an inadequate way of evaluating learning.
- Just because learners PAY ATTENTION doesn’t mean they learned. Measuring attention is an inadequate way of evaluating learning.
- Just because learners ACTIVELY PARTICIPATE in learning doesn’t mean they learned. Measuring participation is inadequate.
- Just because learners ACTIVELY PARTICIPATE in learning doesn’t mean they learned. Measuring participation is inadequate.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
<table>
<thead>
<tr>
<th>Tier</th>
<th>Effects of Transfer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Effects of Transfer: Including outcomes affecting (a) learners, (b) coworkers/family/friends, (c) organization, (d) community, (e) society, and (f) the environs. CERTIFYING EFFECTS OF TRANSFER REQUIRES: Certification of transfer plus a rigorous method of assessing transfer's causal impact—including positive and negative effects.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>Transfer</th>
<th>When learner uses what was learned to perform work tasks successfully—as clearly demonstrated through objective measures.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assisted Transfer—when performance is substantially prompted/supported. ADAPTED TO CERTIFY ASSISTED TRANSFER.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Full Transfer—when learner demonstrates full agency in applying the learning. ADAPTED TO CERTIFY FULL TRANSFER.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>Task Competence</th>
<th>Learner performs relevant realistic actions and decision making.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Task Competence—during or right after learning event.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remembered Task Competence—after several days or more. ADAPTED TO CERTIFY TASK COMPETENCE.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Decision Making Competence</th>
<th>Learner makes decisions given relevant realistic scenarios.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decision Making Competence—during or right after learning event.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remembered Decision Making Competence—after several days or more.</td>
<td></td>
</tr>
</tbody>
</table>

Understanding Remembering
Work Performance

<table>
<thead>
<tr>
<th>Tier</th>
<th>Effects of Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Effects of Transfer: Including outcomes affecting (a) learners, (b) coworkers/family/friends, (c) organization, (d) community, (e) society, and (f) the environ. CERTIFYING EFFECTS OF TRANSFER REQUIRES: Certification of transfer plus a rigorous method of assessing transfer's causal impact—including positive and negative effects.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tier</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>When learner uses what was learned to perform work tasks successfully—as clearly demonstrated through objective measures.</td>
</tr>
<tr>
<td></td>
<td>• Assisted Transfer—when performance is substantially prompted/supported. ADEQUATE TO CERTIFY ASSISTED TRANSFER.</td>
</tr>
<tr>
<td></td>
<td>• Full Transfer—when learner demonstrates full agency in applying the learning. ADEQUATE TO CERTIFY FULL TRANSFER.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tier</th>
<th>Task Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Learner performs relevant realistic actions and decision making.</td>
</tr>
<tr>
<td></td>
<td>• Task Competence—during or right after learning event.</td>
</tr>
<tr>
<td></td>
<td>• Remembered Task Competence—after several days or more. ADEQUATE TO CERTIFY TASK COMPETENCE.</td>
</tr>
<tr>
<td></td>
<td>NOTE: “Tasks” comprise both decision making and action taking. For example, a person learning to write poetry could decide to use metaphor, could opt to use it, or could do both.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tier</th>
<th>Decision Making Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Learner makes decisions given relevant realistic scenarios.</td>
</tr>
<tr>
<td></td>
<td>• Decision Making Competence—during or right after learning event.</td>
</tr>
<tr>
<td></td>
<td>• Remembered Decision Making Competence—after several days or more. ADEQUATE TO CERTIFY DECISION MAKING COMPETENCE.</td>
</tr>
</tbody>
</table>

---

The Learning-Transfer Evaluation Model

data source: Kirkpatrick's Four Success Levels

### Compared to Kirkpatrick-Katzell Four-Level Model of Learning Evaluation

- **Level 4 -- Results**
  - When learners rate what was learned to perform work tasks successfully—on really, measured 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. |

- **Level 3 -- Behavior**
  - 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. |

- **Level 2 -- Learning**
  - 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. |

- **Level 1 -- Reaction**
  - Learner engaged in action related to learning. |
  - Measures of Reaction: |
  - Measures of Interest |
  - Measures of Satisfaction |
  - Measures of Participation |
  - Learner reports that they have completed a learning experience. |

---

The Learning-Transfer Evaluation Model

data source: Kirkpatrick's Four Success Levels

### Kirkpatrick-Katzell Four-Level Model of Learning Evaluation

- **Level 4 -- Results**
  - When learners rate what was learned to perform work tasks successfully—on really, measured 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. |

- **Level 3 -- Behavior**
  - 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. |

- **Level 2 -- Learning**
  - 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. |

- **Level 1 -- Reaction**
  - Learner engaged in action related to learning. |
  - Measures of Reaction: |
  - Measures of Interest |
  - Measures of Satisfaction |
  - Measures of Participation |
  - Learner reports that they have completed a learning experience. |
Two Ways to Use LTEM

1. Assessing Your Evaluations
2. Learning Design & Development: Working Backward from Your Goals
There is no perfect measurement tool!
LTEM is not a panacea!
Do your evaluations with wisdom!!

Of all people’s miseries the bitterest is this, to know so much and to have control over nothing.

Herodotus
(484-425 BC)
These Evaluation Problems

1. Our Most Common Metric (Traditional Smile Sheets) Tell Us Very Little about the Factors of Learning Effectiveness

2. Catastrophic Problems Measuring Learning
   - Bias from Timing of Measurement
   - Bias by Measuring in Same Context
   - Bias by Measuring Knowledge Only

3. Ignore the Role of Remembering

4. Focus Too Much on Meaningless Results
   - Attendance, Participation, Completion

5. We have NOT been able to Make the Best Decisions from Our Learning Evaluations

---

• LTEM Report and Model — https://is.gd/LTEM999
• Katzell’s Contribution — https://is.gd/Katzell
• Updated Smile-Sheet Questions 2018 — https://is.gd/Questions2018
• A Better Net Promoter Question — https://is.gd/replaceNPS
• Be Careful When Benchmarking — https://is.gd/DoNotBenchmark
• Debate About Kirkpatrick Model — https://is.gd/epicbattle
• Better Responses on Smile Sheets — https://is.gd/betterresponses

Subscribe to my Newsletter: — https://www.worklearning.com/sign-up/
Enlightening Ourselves as Learning Professionals

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

Slides available at: www.is.gd/will999stuff