# Knowledge Management & E-Learning

# A customizable evaluation instrument to facilitate comparisons of existing online training programs

**Cheryl A. Murphy**
**Elizabeth A. Keiffer**
University of Arkansas, USA
**Jack A. Neal**
University of Houston, USA
**Philip G. Crandall**
University of Arkansas, USA

# A customizable evaluation instrument to facilitate comparisons of existing online training programs

Cheryl A. Murphy*

College of Education and Health Professions
University of Arkansas, USA
E-mail: cmurphy@uark.edu

Elizabeth A. Keiffer

Department of Mathematical Sciences
University of Arkansas, USA
E-mail: ekeiffe@uark.edu

Jack A. Neal

Conrad N. Hilton College
University of Houston, USA
E-mail: jneal2@uh.edu

Philip G. Crandall

Department of Food Science
University of Arkansas, USA
E-mail: crandal@uark.edu

*Corresponding author

**Abstract:** A proliferation of retail online training materials exists, but often the person in charge of choosing the most appropriate online training materials is not versed in best practices associated with online training. Additionally, the person must consider the context of the training situation when choosing a training solution. To assist this decision-making process an evaluation instrument was developed. The instrument was designed to help decision-makers 1) assess multiple online training programs against known best practices, and 2) consider context specific training needs via a weighting process. Instrument testing across multiple online training programs was performed, and weighted and unweighted results were examined to determine the impact of contextualized weighting. Additionally, evaluation data from the new instrument were compared to data from an existing online training evaluation instrument. Results indicated the new instrument allowed for consistent rankings by raters across multiple programs, and when the new weighting process was applied small differences were magnified making them more noticeable in overall rating scores. Thus the new weighted instrument was effective in 1) assessing multiple online training programs, and 2) providing reviewers clearer context-specific rating data on which they could make purchasing decisions.

**Biographical notes**: Dr. Cheryl A. Murphy is an Associate Professor of Educational Technology in the College of Education and Health Professions at the University of Arkansas. She is the Director of Distance Education for the College and serves as ETEC Program Coordinator. Her research interests include the design, development, and evaluation of online instruction.

Elizabeth A. Keiffer is a Clinical Assistant Professor in the Department of Mathematical Sciences at the University of Arkansas. She is the Assistant Director of the Mathematics Resource and Teaching Center. Her research interests include differential item functioning and instrument/survey creation and evaluation especially in mathematics education.

Dr. Jack (Jay) Neal is an Assistant Professor at the Conrad N. Hilton College of Hotel and Restaurant Management at the University of Houston. He currently manages the Food Production, Sensory and food Microbiology Laboratories. His research interests include food safety behavior, culture and training for non-English speakers.

Dr. Philip G. Crandall is a Professor in the Department of Food Science at the University of Arkansas. His laboratory's research focuses on reducing the risks from Ready-to-Eat foods through employee training, novel cleaning and sanitizing intervention for deli equipment and evaluating antimicrobials.

## 1. Introduction

Training is a universal need that spans all businesses and industries. Employees must be trained on topics ranging from communication skills to proper food safety techniques. Larger operations have training departments and can create their own training materials, but smaller organizations often choose to purchase existing training material. One popular option is the use of e-learning technologies and online training (Bosco, 1986; Neal, Murphy, Crandall, O'Bryan, Keiffer, & Ricke, 2011; Khalili & Shashaani, 1994; O'Bryan, Johnson, Shores-Ellis, Crandall, Marcy, Seideman, & Ricke, 2010; Parlangeli, Marchigiani, & Bagnara, 1999; Strother, 2002). However, a proliferation of online training materials has left decision makers responsible for training in a quandary when choosing which online program to purchase (Barker, 2007; Parker, 2004; Seufert, 2002; Zaied, 2012). This is particularly exacerbated in areas where the decision maker has little or no training background or lacks an understanding of best practices in online training (Strother, 2002).

The person who will be asked to make a decision regarding the adoption of an online training program varies (Barker, 2007). This person may be a training director who has a strong understanding of and dedication to training best practices, a manager who focuses on employee and customer needs, or a Person In Charge (PIC) who concentrates on meeting company needs. Although skill sets within all three of these possibilities overlap, each of these individuals brings unique understandings and knowledge to the decision-making situation. A good online training evaluation tool should combine the best aspects of all of the aforementioned skill areas and make it easy for the decision maker to: 1) Evaluate training relative to known best practices (training director); 2) Assess the training's ability to address the needs of the company (PIC); and

3) Examine the training with employee and customer needs in mind (manager). See Fig. 1 for a pictorial representation of the overlapping skill area an evaluation tool should fill.
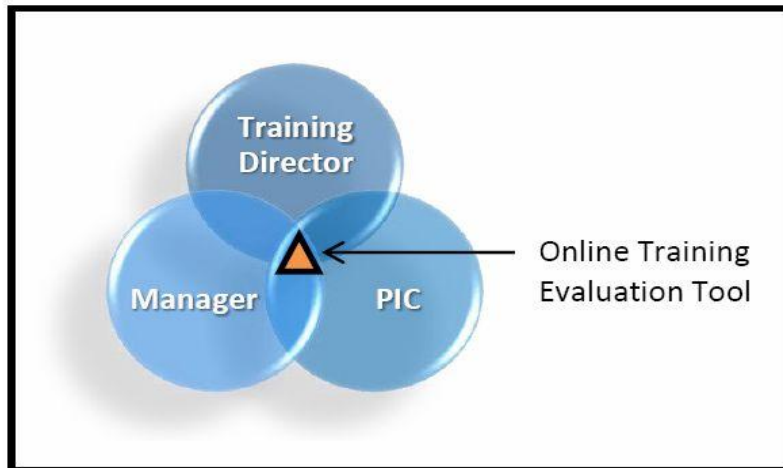


**Fig. 1.** Overlapping skill area filled by evaluation tool

The need for an online training evaluation tool exists in many business and industry contexts, and as previously stated it should perform multiple functions as elearning is a multidimensional construct (Agariya & Singh, 2013). First and foremost, the evaluation tool should assist decision makers in determining if a training program under consideration for purchase adheres to known best practices relative to online training (Zaied, 2012). This includes evaluating areas such as intuitive interface design, logical content sequencing, and appropriate assessment methods. Second, the evaluation tool should determine if the training program under consideration addresses the overarching training needs of the company which often include meeting federal, state, and regional regulations (Strother, 2002). This involves an examination of the training outcomes, content, and activities to ensure all mandated training regulations are met. Third, the evaluation tool should help the decision maker evaluate the online training program in relation to meeting the specific contextual needs of the workplace, employee, and customer (Becker, Fleming, & Keijsers, 2012; Istrate, 2013). This involves a consideration of what aspects are most important within specific business and industry contexts, as each have their own unique needs and demands.

Elliott Masie, founder and president of the MASIE Center, succinctly described the difficulty decision makers have in choosing appropriate contextual training options by stating, "A major challenge that learning professionals are struggling with today is how to place the great abundance of content that is available to them into context for the needs of many different learners" (Skillsoft, 2013). Similarly, Alkhattabi, Neagu, and Cullen (2010) asserted specified contexts and user perspectives must be considered when defining quality in e-learning. The lack of contextually specific training is evidenced in the somewhat generic online training options that are available. While most retail vendors of online training understand the importance of incorporating online best practices and meeting necessary regulations, they are unable to tailor training to meet the needs of all contexts (Becker, Fleming, & Keijsers, 2012; Mosharraf & Taghiyareh, 2013). As a result, most vendors create online training that adheres to best practices while providing content that addresses a broad range of regulations in a general fashion. Therefore, although programs may meet the basic requirements for training, these existing online

training programs may or may not meet the contextual needs of a specific business or industry.

One industry where the need for contextualized training is frequently evidenced but decision makers need assistance in choosing the best training options is food service (Neal et al., 2011; Egan et al., 2007). While many are responsible for food safety along the farm to fork continuum, the responsibility for food safety at the food production level often rests with PIC's, while training responsibilities at the retail food service level typically lie with food service managers (USFDA, 2000; USFDA, 2004; Neal et al., 2011). Multiple online food safety training programs are available to assist food service decision makers. However, the challenge for these individuals is evaluating programs to determine which, if any, are instructionally sound (Egan et al., 2007) and also meet the contextual needs of the organization. This is no small task given many training decision makers in food service have little or no instructional experience and are not experts at evaluating online learning materials or environments.

In addition to evaluating online training programs for adoption decision makers must also consider the evaluation of training programs after adoption (Egan et al., 2007; Peak & Berge, 2007; Strother, 2002). As well as providing proactive guidance during the initial adoption of a training program an online training evaluation tool could be used reactively to compare newly developed products to existing training programs, address changes that impact the return on investment (ROI), or address changes in regulations or technologies that impact online training. A solid online training evaluation tool would allow the decision maker to be both proactive and reactive to the training demands of the organization.

Despite the potential lack of experience evaluating online programs, achieving a thorough evaluation of online training options is critical for the decision maker. The evaluation should result in comparative data the decision maker can use to make informed choices in selecting suitable training materials. The development of a tool that can be used by decision makers to evaluate multiple online programs for best practices, meeting content needs, and contextual appropriateness is a solid step. Such an instrument should be comprehensive yet provide the necessary information to meet the decision maker's needs with minimal time commitment (Neal et al., 2011). The remainder of this article details the development and testing of such an instrument.

## 2.   Instrument development

The development of an online instrument that assists decision makers in the evaluation of multiple online programs relative to best practices, meeting content needs, and contextual appropriateness was a three-step process. During the first step the authors tested an existing instrument to determine the effectiveness of the instrument in performing the aforementioned tasks. Results of this action were used in the second step to inform a major revision of the original instrument, which was then tested for content validity. The third step in the development of the instrument involved the inclusion of a Delphi panel to inform a customization process which added weighting to the revised instrument.

As previously mentioned, many businesses and industries need assistance in the decision-making process when determining which existing online training to adopt. Because of the strong need for contextualized evaluation of online training relative to the food service industry, all instrument analysis was performed within the context of this setting. More specifically, online training modules concerning food safety were used in

the testing of the instrument. Details regarding all three steps as well as the completed instrument are provided below.

## 2.1. Testing an existing instrument

In 1997 Ginger Pisik developed a form that could be used by managers to evaluate online training. Her 68-item instrument assessed five areas including: Content; Learners; Job transfer; Design and packaging; and Operation. This instrument was slightly modified by Neal et al. (2011) to delete references to obsolete technologies. The 68-item revised instrument was tested in a food service context to determine its ability to effectively assess existing online training. Persons in Charge used the modified instrument to evaluate online food safety training modules. The objective of the study was to provide the PIC an assessment tool that could identify strengths, weaknesses, and usability of current retail food safety programs.

Three online food safety training modules were evaluated and subjects indicated the degree to which they were able to assess and evaluate the online food safety modules with the modified instrument. Data were also examined to determine consistency within instrument responses across the training modules. The results indicated the modified instrument did allow for numeric comparisons on best practice aspects of the three online training modules. However, data also reflected that even though respondents reported the instrument was too lengthy the instrument was not detailed enough to provide a thorough assessment of the online training relative to the specific needs of a retail food service setting (Neal et al., 2011). In other words, the instrument proved effective in comparing online training programs relative to best practices in online training, but not in relationship to specific contextual needs of the job or work area. Additionally, respondents indicated confusion relative to some items due to unfamiliar terminology or unclear wording. The findings of this first study demonstrated the need for a revised evaluation instrument that is: 1) based on general best practices of online training (such as the Pisik instrument); 2) less time-consuming; 3) clearly worded with familiar terms; and 4) capable of being tailored to assess an online program's ability to meet the specific needs of various contexts such as those found in retail food service.

## 2.2. Instrument revision

For the second study the modified Pisik instrument used in Neal et al. (2011) was revised. To address shortcomings identified in the first study this instrument revision was more extensive than previous changes. The authors incorporated the most recent instructional design wisdoms for online learning to ensure the evaluation of general best practices. Redundancies within question items were identified and eliminated to streamline the instrument. Additionally, language and wording were altered to be more common and reflective of online *training* as opposed to general online learning.

More specifically, to create the new instrument the original Pisik (1997) categories (5) and items (68) were updated to incorporate critical online design components identified within newer online evaluation tools. These tools included an updated version of the Pisik instrument (Pisik, 2004), Quality Matters Rubric Standards (MarylandOnline, 2011), Quality Standards for Evaluating Multimedia and Online Training (Gillis, 2000), and the Quality Online Course Initiative (Illinois Online Network, 2010). To accomplish the update the aforementioned instruments were examined and key constructs as well as sub-constructs were identified across all instruments. Constructs and sub-constructs that appeared in most instruments but not in the original Pisik were

considered for inclusion in the new instrument. Redundancies were removed to streamline the instrument as much as possible and items were added or altered to more accurately reflect the key constructs and sub-constructs. Items from the original instrument that reflected outdated technologies or practices were removed. Lastly, the resulting items were grouped into categories reflective of current training language and terminology. The result was the creation of the Customizable Tool for Online Training Evaluation (CTOTE) which includes four categories and forty-eight survey items. A comparison of the categories and number of items amongst the original Pisik instrument (1997), the modified Pisik (Neal et. al, 2011), and the Customizable Tool for Online Training Evaluation (CTOTE) can be seen in Table 1.

**Table 1**
A comparison of categories and items across evaluation instruments

| Instrument | Categories | Number of Items |
|---|---|---|
| Original Pisik (1997) | Content and Instruction | 38 |
| | Learners | 6 |
| | Job Transfer | 5 |
| | Design and Packaging | 8 |
| | Operation | 11 |
| | Total Items: | 68 |
| | | |
| Modified Pisik (Neal et. al. 2011) | Content and Instruction | 38 |
| | Learner Interaction | 6 |
| | Transferability of Knowledge | 5 |
| | Design and Packaging | 8 |
| | Operation | 11 |
| | Total Items: | 68 |
| | | |
| CTOTE (Murphy et al., 2012) | Content and Outcomes | 17 |
| | Assessment and Transfer | 7 |
| | Technology Design and Requirements | 12 |
| | Operations and Support | 9 |
| | Total Items: | 45 |

To ensure the content validity of the new items and categories the CTOTE instrument was tested using the method prescribed by the Indexes of Item-Objective Congruence for Multidimensional Items (Turner & Carlson, 2003). This method incorporates the use of content experts to assess the extent to which items on an instrument accurately measure the specific objectives (categories in this instance) under

which they are listed. Instructional designers (n=6) with expertise in online training assessed all items from the CTOTE instrument for content validity using the IIOC method. Thirteen items were identified as being of interest because their corresponding IIOC value fell below 0.67, a value chosen to represent agreement between 4 of the 6 instructional designers who completed the questionnaire. The items were revised and results indicated consensus among the experts that all items within the revised instrument were accurate measures of the categories in which they were grouped.

A reliability study was performed to ensure the CTOTE instrument provided consistent evaluation data. Cronbach Alpha scores of 0.83 and above were calculated for all scales, exceeding the standard acceptable reliability coefficient of 0.7 (Nunnaly, 1978). These results indicated the CTOTE instrument was able to provide consistent evaluation responses. Reliability calculation results are presented in Table 2.

**Table 2**
Reliability calculation results for the CTOTE instrument

| Scale | N | Number of Items | Mean | SD | Cronbach Alpha | SEM |
|---|---|---|---|---|---|---|
| Content and Outcomes | 188 | 17 | 83.91 | 13.07 | 0.94 | 3.20 |
| Assessment and Transfer | 193 | 7 | 80.15 | 14.36 | 0.83 | 5.92 |
| Technology Design and Requirements | 192 | 12 | 81.44 | 12.94 | 0.84 | 5.18 |
| Operations and Support | 195 | 9 | 81.19 | 14.25 | 0.85 | 5.52 |

*Total sample size reliability study was 195. Only scale responses with complete data were used when calculating descriptive statistics.

Thus, the construction of the CTOTE instrument was the second step in the process of creating an evaluation instrument grounded in best practices of online training and capable of meeting specific needs of various contexts.

## 2.3. Customization process

The final step in the instrument development process was the incorporation of a method whereby the instrument could be used to meet the needs of various contexts. To accomplish this task items within the CTOTE instrument were weighted based on contextual importance. To determine contextual importance input was sought from an expert panel relative to the context in which the online training was being sought. As an example, for our specific study we focused on food service and in particular food safety online training. Therefore, the expert panel consisted of twelve experienced retail food service managers who were familiar with food safety training. To use this process in different contexts, the panel of experts would change depending on the context for which the training was being examined, and the number of experts could vary from five to fifteen.

Once panel members were identified the expert panel used a Delphi process to rank items on the CTOTE instrument they viewed as most important, somewhat important, and mildly important when considering the purchase of online training to meet their particular needs. Again, relative to our study this involved online training for retail food service settings, so our experts rated each item on the CTOTE in relation to its importance in food safety training in retail food service. The Delphi process was used until consensus was garnered across the expert panel regarding the importance level of each item. The results of this expert panel Delphi process informed the weighting of items within the CTOTE instrument, with more weight given to the evaluation areas deemed by the expert panel as most important (3) and less weight given to those deemed somewhat important (2) and mildly important (1).

By including the Delphi process with an expert panel the result was an evaluation instrument that provided an overall evaluation score for an online training module that was based on best practices of online training as well as the most important elements relative to training in in a specific context as determined by experts from within that context. In essence, the expert-driven weighting provided an online training evaluation instrument that offered numeric data that could be used in cross-program comparisons, and was also tailored to the specific needs within a particular context.

As an example of the effect of this contextualized weighting, Fig. 2 and 3 demonstrate how overall scores within a category were impacted when weighting is applied. In the fictitious example provided in Fig. 2 no weighting was applied and the CTOTE instrument was scored without the benefit of the contextualized Delphi input. The end result was a category score of 75%.

| Item | SA | A | DA | SD | N/A | Total | Total Possible | Total NA (Total Possible for NA item) |
|------|----|---|----|----|-----|-------|----------------|----------------------------------------|
| 1 | 4 | 3 | 2 | 1 | 0 | 0 | 4 | 4 |
| 2 | 4 | 3 | 2 | 1 | 0 | 4 | 4 | |
| 3 | 4 | 3 | 2 | 1 | 0 | 1 | 4 | |
| 4 | 4 | 3 | 2 | 1 | 0 | 3 | 4 | |
| 5 | 4 | 3 | 2 | 1 | 0 | 4 | 4 | |
| | | | | | Totals= | 12 | 20 | 4 |
| | UNWEIGHTED CATEGORY PERCENTAGE [Total/(Total Possible overall – Total NA)] x 100 | | | | | | [12/(20 – 4)] x 100 = **75%** | |

**Fig. 2.** Unweighted example of CTOTE instrument scoring

However, when weighting was applied to the same fictitious example in Fig. 3 the end result changed. Each item in Fig. 3 was assigned a weight based on results of the Delphi panel. Items 1 and 2 were given higher weight than item 4, and the lowest weighted items were 3 and 5. As demonstrated in these examples, adding weights impacted the Category Percentage. In the examples items 2 and 5 were ranked the same, but when weights were added in Fig. 3 item 2 became much more important to the overall category total, increasing the Category Percentage from 75% to 82%.

| Item | SA | A | DA | SD | N/A | Weight | Total (rating x weight) | Total Possible (max rating x weight) | Total NA (Total Possible for NA item) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 3 | 2 | 1 | (0) | x3 | 0 | (4x3)= 12 | 12 |
| 2 | (4) | 3 | 2 | 1 | 0 | x3 | 12 | (4x3)= 12 | |
| 3 | 4 | 3 | 2 | (1) | 0 | x1 | 1 | (4x1)= 4 | |
| 4 | 4 | (3) | 2 | 1 | 0 | x2 | 6 | (4x2)= 8 | |
| 5 | (4) | 3 | 2 | 1 | 0 | x1 | 4 | (4x1)= 4 | |
| | | | | | | Totals = | 23 | 40 | 12 |
| | **WEIGHTED CATEGORY PERCENTAGE** [Total/(Total Possible overall – Total NA)] x 100 | | | | | | [23/(40 − 12)] x 100 = **82%** | | |

**Fig. 3.** Weighted example of CTOTE instrument scoring

This weighting system (again, derived from experts who worked in the specific context where the training will be used) offered the ability to "customize" the instrument to meet the needs of particular learners and the specific organization. This offered the decision maker more precise and applicable information when evaluating multiple online programs.

The Delphi weighting method was utilized and tested in the authors' second study. As previously mentioned, twelve experienced retail food service managers familiar with food safety training participated in the Delphi process. The resulting weights can be viewed in the complete CTOTE instrument in Fig. 4a, 4b, 4c, 4d (see Appendix).

## 2.4. Instrument calculations

As noted in Fig. 4, to calculate the weighted percentage for a category you must take into account the total points accumulated in that category, but you must also factor in the total points possible as well as the number of Not Applicable answers that were provided. To calculate the Total points multiply the reviewer ranking (in this case 0-4) by the weight, and add the results to get the Total points for the category.

To calculate the Total Possible points multiply the maximum ranking possible (in this case 4) by the weight for each item. Add the results to get the Total Possible Points for the category. Lastly, to account for all Not Applicable questions and remove them so they are not influencing the overall score deduct the Total Possible scores for all items rated as NA from the Total Possible.

The final step in calculating the Category Percentage is to divide the Total points received from the Total Possible (minus the NAs), and multiply by 100. This will provide a weighted percentage for this particular category that can then be compared to the same weighted percentage for this category when reviewing additional online training programs. To calculate the percentage score for all four CTOTE categories add the Category Percentages for all four categories together and divide by four. As with the Category Percentages this overall percentage score can be used to compare multiple online training programs.

## 3.   Instrument testing

Once the instrument and weightings were devised the CTOTE instrument was tested. Content and face validity were established by the IIOC results from professional online trainers and the expert Delphi panel, both of which were described above. Testing across instruments was performed to examine the CTOTE instrument in relationship to the original Pisik (1997) instrument and to examine weighted versus unweighted results to determine the impact on ratings as a result of the weighting procedure. Lastly, testing was performed across programs to ensure CTOTE provided consistent evaluation data across multiple training programs and modules. Details regarding these tests are provided below.

### 3.1. Testing across instruments

To test the CTOTE instrument in relation to the Pisik (1997) instrument a retail online food safety training program was examined by twenty-six reviewers (*N*=26). Reviewers were instructed to use both the Pisik instrument and the CTOTE instrument to evaluate the online training program. Data in some areas were not recorded, but where all values were present scores, means, and standard deviations were calculated for each instrument. Correlations were then calculated using complete matched data. This determined if the results from the Pisik (1997) instrument and the CTOTE instrument (weighted and not weighted) were reporting the same conclusions.

Correlations between the weighted and unweighted versions of our instrument were very strong and positive, $r(23) = .90$, $p < .05$. Correlations between the unweighted CTOTE instrument and the Pisik instrument were also positive and were statistically significant $r(18) = .55$, $p < .05$, but not as strong. In fact, the weighted instrument versus the Pisik had the lowest (albeit still significant) positive correlation at $r(19) = .47$, $p < .05$. This indicates all instruments were reporting *similar* conclusions, but the difference between the weighted version and the Pisik was greater than the unweighted when correlated with the Pisik instrument. Statistical results of the correlations including means, standard deviations, and correlation coefficients may be viewed in Table 3.

**Table 3**
Correlations between CTOTE unweighted, weighted, and Pisik evaluation results

| Instrument and Comparisons | n | Mean | Standard Deviation | Correlation Coefficient |
|---|---|---|---|---|
| CTOTE - no weight | 25 | 82.33 | 10.82 | |
| CTOTE - weighted | 26 | 77.61* | 14.38 | |
| Pisik 1997 | 22 | 87.12* | 11.14 | |
| **Correlations** | | | | |
| CTOTE –no weight vs. weighted | 25 | | | r = .90* |
| CTOTE no weight vs. Pisik | 20 | | | r = .55* |
| CTOTE weighted vs. Pisik | 21 | | | r = .47* |

*\* = statistically significant at alpha=.05*

### 3.2. Testing across programs

To test how the CTOTE instrument would perform when examining multiple programs four separate retail online food safety training programs covering the same general content were examined by forty-eight reviewers (*N*=48). Each reviewer compared two of the programs and used the CTOTE weighted instrument to evaluate both, producing a total of ninety-six CTOTE reviews. See Table 4 for a breakdown of reviewer assignments.

**Table 4**
Breakdown of reviewer assignments

| Review Groups | Retail Online Food Safety Training Programs | | | | |
| | Program 1 | Program 2 | Program 3 | Program 4 | *Total Responses* |
| --- | --- | --- | --- | --- | --- |
| Review Group 1 | 8 | 8 | | | *16* |
| Review Group 2 | 8 | | 8 | | *16* |
| Review Group 3 | 8 | | | 8 | *16* |
| Review Group 4 | | 8 | 8 | | *16* |
| Review Group 5 | | 8 | | 8 | *16* |
| Review Group 6 | | | 8 | 8 | *16* |
| *Total Reviews* | *24* | *24* | *24* | *24* | *96* |

Overall CTOTE percentage scores from individual reviews were calculated and combined with the overall percentage scores across each program and divided by the total number to get the overall percentage rating score for each program. Based on overall ratings from the CTOTE instrument preferred programs were identified and can be seen in Table 5. Data indicated the programs were ranked consistently by raters, with Program 1 and Program 2 consistently ranked higher than Program 3 or Program 4. Program 1 was the highest ranking program while Program 4 was clearly the lowest ranking program.

## 4. Results

As indicated above, results between the weighted and unweighted versions of the CTOTE instrument were strong and positive. Results between the CTOTE instrument (weighted and unweighted) and the Pisik (1997) instrument were also positive and were statistically significant, but not as strong. These results demonstrated when rating the same program using the two instruments the overall scores were lowest for the weighted CTOTE instrument and highest for the Pisik (1997) instrument. This difference was statistically significant and demonstrates that when weighting is applied the ratings are more extreme because a wider range is used. In other words, it is more difficult to get a high score if the program falls short in an important area, and easier to get a high score if the program

excels in important areas. In both instances the weighting makes strengths or weakness in important areas more noticeable in the overall percentage score. These findings support the use of weighting to allow for more scrutinous contextual evaluation, thus providing the reviewer more precise data on which to make purchasing decisions relative to online training.

**Table 5**
Consistency of overall rating outcomes for the CTOTE across multiple programs

| Training Program Comparisons | n | Favored Program 1 | Favored Program 2 | Favored Program 3 | Favored Program 4 |
|---|---|---|---|---|---|
| Program 1 vs. Program 2 | 10* | 5 | 5 | | |
| Program 1 vs. Program 3 | 8 | 6 | | 2 | |
| Program 1 vs. Program 4 | 8 | 6 | | | 2 |
| Program 2 vs. Program 3 | 8 | | 6 | 2 | |
| Program 2 vs. Program 4 | 7 | | 4 | | 3 |
| Program 3 vs. Program 4 | 7 | | | 7 | 0 |
| Totals | 48 | 17 | 15 | 11 | 5 |

\* = two reviewers mistakenly reviewed program 1 vs. 2 instead of 2 vs. 4, and 3 vs. 4

Results also demonstrated the weighting process helps reviewers make more precise distinctions based on items that are deemed most important to their circumstance by experts (ie. the Delphi panel). As previously stated, this method makes shortcomings in the areas of most contextual importance more apparent because if the program falls short in an area deemed highly important it is more noticeable in the overall score. Thus results are more contextualized, allowing for informed evaluation and purchasing decisions relative to context.

Lastly, data indicated the programs were ranked consistently by raters across multiple programs. The consistency in rankings across raters and programs demonstrated that the CTOTE instrument may be used to rank multiple programs, thus allowing a PIC, trainer, or manager to evaluate and compare multiple retail online training options.

## 5.   Conclusion

As indicated by researchers (Barker, 2007; Parker, 2004; Seufert, 2002; Zaied, 2012) it is difficult for companies to decide which eLearning provider or vendor to choose. This research sought to develop and test an instrument that would assist uninformed decision-makers in the evaluation of multiple retail online training programs. The instrument was developed to take into consideration known best practices relative to online training (Zaied, 2012), overarching training needs relative to content (Strother, 2002), and the specific contextual needs of the workplace, employee, and customer (Becker, Fleming, & Keijsers, 2012; Istrate, 2013). The instrument was validated and tested within the area of food service, an industry where need for contextualized training is evidenced but decision makers may be ill-equipped to choose the best training options (Neal et al., 2011; Egan et al., 2007).

Instrument testing across multiple online food safety training programs produced consistent rankings amongst raters, demonstrating the consistency of the evaluation instrument. Similarly, when comparing ratings from the CTOTE instrument (weighted and unweighted) with an existing evaluation instrument (Pisik, 1997) positive correlations indicated all evaluation three instrument versions rated the programs in a similar manner. Again, this demonstrates consistency of the instrument. However, correlation results also indicated the weighted CTOTE instrument facilitated a larger percentage score range than the unweighted CTOTE or the Pisik (1997) instrument. This indicated the use of an expert Delphi panel to inform weighting was effective in accentuating small but potentially critical contextual differences between online food safety training programs. These differences were magnified by the weighting process, making them more noticeable in the overall percentage scores. Thus, the results of this research indicated the weighted CTOTE instrument ultimately provided the reviewer clearer data on which decision to make relative to the purchase of retail online food safety training.

## 6. Future studies

This research has been undertaken in the context of the retail food service industry. Although the incorporation of the Delphi process is designed to account for contextual differences, additional studies should examine the use of the CTOTE instrument in relation to other industries to support and further strengthen the findings of this study.

## Acknowledgements

## References

Agariya, A. K., & Singh, D. (2013). e-Learning quality: Scale development and validation in Indian context. *Knowledge Management & E-Learning (KM&EL)*, 4(4), 500–517.

Alkhattabi, M., Neagu, D., & Cullen, A. (2010). Information quality framework for e-learning systems. *Knowledge Management & E-Learning (KM&EL), 2*(4), 340–362.

Barker, K. C. (2007). E-learning quality standards for consumer protection and consumer confidence: A Canadian case study in e-learning quality assurance. *Educational Technology & Society, 10*(2), 109–119.

Becker, K., Fleming, J., & Keijsers, W. (2012). E-learning: Ageing workforce versus technology-savvy generation. *Education + Training, 54*(5), 385–400.

Bosco, J. (1986). An analysis of evaluations of interactive video. *Educational Technology, 26*(5), 7–17.

Egan, M. B., Raats, M. M., Grubb, S. M., Eves, A., Lumbers, M. L., Dean, M. S., & Adams, M. R. (2007). A review of food safety and food hygiene training studies in the commercial sector. *Food Control, 18*, 1180–1190.

Gillis, L. (2000). *Quality standards for evaluating multimedia and online training*. Whitby, Canada: McGraw-Hill Ryerson.

Illinois Online Network. (2010). ION's quality online course initiative. Retrieved from http://www.ion.uillinois.edu/initiatives/qoci/

Istrate, O. (2013). Criteria for elearning programmes evaluation. *Journal of Advanced Distributed Learning Technology, 1*(1), 27–35.

Khalili, A., & Shashaani, L. (1994). The effectiveness of computer applications: A meta-analysis. *Journal of Research on Computing in Education, 27*(1), 48–61.

MarylandOnline. (2011). *Quality Matters rubric standards 2011 – 2013 edition with assigned point values*. Retrieved from http://www.qmprogram.org/rubric

Mosharraf, M., & Taghiyareh, F. (2013). Qualitative development of eLearning environments through a learner relationship management methodology. *Knowledge Management & E-Learning (KM&EL), 5*(1), 56–65.

Murphy, C., Keiffer, E., Neal, J., & Crandall, P. (2012). *A customized evaluation instrument to facilitate comparisons across online training programs: CTOTE*. Paper presented at the Annual Association of Educational Communications and Technology Conference, Louisville, KY, USA.

Neal, J. A., Murphy, C. A., Crandall, P. G., O'Bryan, C. A., Keiffer, E., & Ricke, S. C. (2011). Development of an evaluation tool for online food safety training programs. *Journal of Food Science Education, 10*(1), 9–12.

Nunnaly, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

O'Bryan, C., Johnson, D. M., Shores-Ellis, K., Crandall, P., Marcy, J., Seideman, S., Ricke, S. (2010). Designing an affordable usability test for e-learning modules. *Journal of Food Science Education, 9*(1), 6–10.

Parker, N. K. (2004). The quality dilemma in online education. In T. Anderson & F. Elloumi (Eds.), *Theory and Practice of Online Learning* (pp. 385 – 421). Retrieved from http://www.aupress.ca/index.php/books/120146

Parlangeli, O., Marchigiani, E., & Bagnara, S. (1999). Multimedia systems in distance education: effects of usability on learning. *Interacting with Computers, 12*(1), 37–49.

Peak, D., & Berge, Z. L. (2006). Evaluation and elearning. *Turkish Online Journal of Distance Education, 7*(1), 124–131.

Pisik, G. B. (1997). Is this course instructional sound? A guide to evaluating online training courses. *Educational Technology, 37*(4), 50–59.

Pisik, G. B. (2004). Is this course instructionally sound? An updated guide to evaluating online training courses. *Educational Technology, 44*(1), 57–60.

Skillsoft. (2013). *From boardroom to desktop: Skillsoft connects corporate strategy to the performance needs of individual learners with new Enterprise Learning Connection* [Press Release]. Retrieved from http://www.skillsoft-usa.net/about/press_room/press_releases/may_02_06_elc.asp

Strother, J. B. (2002). An assessment of the effectiveness of e-learning in corporate training programs. *The International Review of Research in Open and Distance Learning, 3*(1).

Seufert, S. (2002). eLearning business models: Strategies, success factors, and best practice examples. In C. Wankel & R. DeFillippi (Eds.), *Rethinking Management Education for the 21st Century* (pp. 109–132). Greenwich, CT: Information Age Publishing.

Turner, R. C., & Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing, 3*(2), 163–171.

United States Food and Drug Administration (USFDA). (2004). *Report of the FDA retail food program database of foodborne illness risk factors*.

United States Food and Drug Administration (USFDA). (2000). *FDA report on the occurrence of foodborne illness risk factors in selected institutional foodservice, restaurant, and retail food store facility types*.

Zaied, A. N. H. (2012). Multi-criteria evaluation approach for e-learning technologies: Selection criteria using AHP. *International Journal on E-Learning, 11*(4), 465–485.

**Appendix**

| Content and Outcomes Category | SA | A | DA | SD | N/A | Weight | Total (rating x weight) | Total Possible (max rating x weight) | Total NA (Total Possible for NA item) |
|---|---|---|---|---|---|---|---|---|---|
| Training outcomes are explicitly stated. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Training outcomes are clear, comprehensive and measurable. | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Training outcomes match training needs. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Content is accurate and current. | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Level of difficulty is appropriate for trainees (i.e. reading-level, writing ability, computer literacy, vocabulary, terminology, and content). | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Content is free from discriminatory examples, terminology, and negative stereotyping (e.g., cultural, racial, gender). | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Outcomes are presented at the beginning of each module/lesson. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Information is "chunked" or grouped into small and convenient pieces to help trainees flexibly learn content. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Content is sequenced in a logical and appropriate manner to aide in trainee learning. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Training program engages trainee. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Examples or exercises are appropriate and relevant. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Examples or exercises are clear, logical, and accurate. | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Examples or exercises are sufficient in number and variety. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Lessons/Exercises help trainee master training program outcomes. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program requires trainee to be an active participant. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Training program offers a variety of instructional delivery methods to accommodate individual learning styles. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program explains new material by expanding on material either from previous knowledge or from the trainee's job experience. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| | | | | | | Totals = | *x* | **124** | *y* |
| WEIGHTED CONTENT AND OUTCOMES PERCENTAGE [Total/(Total Possible overall – Total NA)] x 100 | | | | | | | | $[x/(124 - y)]$ x 100 = _____ % (C & O %) | |

**Fig. 4a.** CTOTE instrument with expert food safety Delphi weightings

| Assessment and Transfer Category | SA | A | DA | SD | N/A | Weight | Total (rating x weight) | Total Possible (max rating x weight) | Total NA (Total Possible for NA item) |
|---|---|---|---|---|---|---|---|---|---|
| Material is presented in realistic job scenarios that can easily be transferred when trainee is back on the job. | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Training program provides a pre-test to determine trainee's readiness for the course. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program provides "self-checks" to assess understanding of content. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Assessment questions measure stated training program outcomes. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program provides performance summary statistics to trainees. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Trainees are provided feedback during lesson exercises. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Feedback for incorrect answers provides additional information, further explanation, and/or an opportunity to try again. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| | | | | | | Totals = | $x$ | 48 | $y$ |
| WEIGHTED ASSESSMENT AND TRANSFER PERCENTAGE [Total/(Total Possible overall – Total NA)] x 100 | | | | | | | | $[x/(48 - y)] \times 100 = $ _____ % (A & T %) | |

**Fig. 4b.** CTOTE instrument with expert food safety Delphi weightings

| Technology Design and Requirements Category | SA | A | DA | SD | N/A | Weight | Total (rating x weight) | Total Possible (max rating x weight) | Total NA (Total Possible for NA item) |
|---|---|---|---|---|---|---|---|---|---|
| Technical requirements can be met with existing hardware/networking. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program is easy to install or access online. | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Interface design and navigation is logical, consistent, and efficient throughout training program (e.g., icon locations, keys used, menu terms, screen layout). | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Navigation features are easy to identify and understand throughout training program (e.g., links, buttons, icons, keys). | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| The layout of on-screen text is clear and readable (including font type, size, and color). | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| The visual layout of the content is well-organized. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program effectively orients trainees to their location within the program (e.g. screen titles, directional icons, menus, site maps). | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program provides links that are self-describing (i.e. clearly explains where the link will go and what will be there). | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Media is integrated into the training program and enhances learning. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Media is relevant and not a distraction. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Trainees can save their work, exit the program, come back to the program at a later time, and start working where they left off. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| The training program is compliant with ADA (American Disabilities Act) standards. | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| | | | | | | Totals = | *x* | **76** | *y* |
| WEIGHTED TECH DESIGN AND REQUIREMENTS PERCENTAGE [Total/(Total Possible overall – Total NA)] x 100 | | | | | | | $[x/(76 - y)]$ x 100 = _____ % (TD & R %) | | |

**Fig. 4c.** CTOTE instrument with expert food safety Delphi weightings

| Operations and Support Category | SA | A | DA | SD | N/A | Weight | Total (rating x weight) | Total Possible (max rating x weight) | Total NA (Total Possible for NA item) |
|---|---|---|---|---|---|---|---|---|---|
| Directions are clear and easy to understand (jargon-free). | 4 | 3 | 2 | 1 | 0 | x3 | | (4x3)= 12 | |
| Directions explain how to find and proceed through various training program components. | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| Directions for exercises make clear where to start and what steps to follow. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Help features are useful and easy to understand. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Help features are specific to instructional material. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Help features can be accessed at any time. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program allows trainees to control the speed at which they proceed through the program. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| A glossary of key terms is easy to access. | 4 | 3 | 2 | 1 | 0 | x1 | | (4x1)= 4 | |
| Training program lets trainees' print important information (e.g., reference pages, lesson summaries). | 4 | 3 | 2 | 1 | 0 | x2 | | (4x2)= 8 | |
| | | | | | | Totals = | *x* | **52** | *y* |
| **WEIGHTED OPERATIONS AND SUPPORT PERCENTAGE** [Total/(Total Possible overall – Total NA)] x 100 | | | | | | | | $[x/(52 - y)]$ x 100 = _____ % (O & S %) | |

**Fig. 4d.** CTOTE instrument with expert food safety Delphi weightings