Customizable Tool for Online Training Evaluation

Cheryl A. Murphy, Elizabeth A. Keiffer, Jack A. Neal, and Jessica Howton

Contents

C.A. Murphy (✉)
Educational Technology, College of Education and Health Professions, Univeristy of Arkansas, Fayetteville, AR, USA
e-mail: cmurphy@uark.edu

E.A. Keiffer
Mathematical Sciences, J. William Fulbright College of Arts and Sciences, University of Arkansas, Fayetteville, AR, USA
e-mail: ekeiffe@uark.edu

J.A. Neal
Conrad N. Hilton College of Hotel and Restaurant Manegement, University of Houston, Houston, TX, USA
e-mail: jneal2@uh.edu

J. Howton
H-E-B Grocery Company, LP, San Antonio, TX, USA
e-mail: jesshowton@gmail.com

J. M. Spector et al. (eds.), *Learning, Design, and Technology*,
DOI 10.1007/978-3-319-17727-4_22-1

Abstract

A proliferation of retail online training materials exists for many industries, but often the person in charge of choosing the most appropriate online training materials is neither a training expert nor versed in best practices associated with online training. To this end, it is critical that uninformed decision makers have access to an easy-to-use evaluation tool which allows for the assessment of strengths and weaknesses among multiple online training programs. Additionally, this tool must take into account the context of the training situation to ensure the chosen program is not only instructionally sound but also meets contextually specific training needs. This article describes the creation, testing, and application of the Customizable Tool for Online Training Evaluation (CTOTE), an evaluation instrument developed 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. The three-step development process is explained including item selection and revision, determination of content validity and reliability, and the use of a Delphi panel to inform contextualized weighting. The instrument is then tested across multiple online training programs with results compared to an established online training evaluation instrument to illustrate the impact of the contextualized weighting. Lastly, the application of the instrument in a specific industry setting (food service) is presented to demonstrate the effectiveness of the instrument in this setting and to establish the potential of the CTOTE in helping uninformed decision makers assess multiple online training programs and make effective context-specific purchasing decisions.

Keywords

Evaluation instrument • Online training • Business and industry • Customizable • Contextual

Introduction

The need for training in business and industry environments is universal and ongoing. All organizations must train new employees on topics ranging from communication skills to proper safety techniques to ensure workers possess the professional competencies and skills needed to safely and successfully perform required job duties. Similarly, skill updates are a necessity as worker knowledge can become obsolete in 3–5 years due to factors such as new or revised regulations, shifting markets, and technological advancements (Steele, 2015; Zhang & Nunamaker, 2003). The ever-changing duties and expectations placed on employees makes continuing education and retraining a necessary activity of today's modern worker (Caudill & Reeves, 2014), and the provision of this training is the responsibility of every business, industry, and organization who wishes to remain competitive and competent.

Recognition of the importance of training is demonstrated by businesses and organizations via the investments they are willing to make in employee professional development. Van Rooij (2011) points out that even as businesses experience economic downturns, they continue to invest in employee training, and the *2014 Industry Report* indicates that training expenditures increased by 11.7 % from the previous year with over $61.8 billion spent in 2014 alone (Training, 2014). A key component of the increase in training expenditures is the growth in the use of technology and online instruction (also referred to as e-learning) to deliver training.

Online Training

According to the Association for Talent Development's most recent annual report, 38 % of training is delivered using technology, and 25 % of total training hours are completed online. Although the presence of an instructor across all training modalities is predominant, 30 % of training is performed without an instructor present and 16 % of all reported training is performed using self-paced online programs (ATD Research, 2015). These numbers are supported by recent eLearning Industry findings which cite that 28.5 % of all training hours in 2014 were delivered online or on a computer without the presence of an instructor (Pappas, 2015).

As the data above indicate, companies are investing in employee training, and they are seeking alternatives to traditional training in an effort to keep pace with growing training needs in an affordable and accessible manner (Steele, 2015). Caudill and Reeves (2014) suggest the cost-effectiveness of online training solutions makes them more attractive to businesses than more expensive traditional methods. Similarly, Batalla-Busquets and Pacheco-Bernal (2013) assert that corporations and trade industries benefit from online training options because workers are able to easily access and use online offerings to continuously update skills. Accordingly, the cost and accessible nature of online training has led many to believe that online training options can give organizations a competitive edge (Jan, Lu, & Chou, 2012), which has led to increased interest and use of this training delivery option. As interest and use has increased, so have the number of commercial online training providers.

Providers of Online Training

In larger organizations, the provision of training often emanates from training departments that create instruction, including online training, to meet the specific needs of their employees. However, the production of high-quality e-learning materials can be costly, and it can be financially more efficient to purchase existing online training materials as opposed to creating them in-house. Smaller organizations, on the other hand, do not have in-house training departments and are, by default, relegated to purchasing commercial training materials to provide employees with professional development opportunities (Neal, Murphy, Crandall, O'Bryan, Keiffer, & Ricke, 2011). Both of these circumstances, paired with the general increase of interest in online training, have contributed to a burgeoning market for online training providers.

As indicated in a recent Forbes magazine article authored by T. J. McCue:

> Online learning, also known as e-learning, is booming. Market research firm Global Industry Analysts projects it will reach $107 Billion in 2015. More traditional methods of training or education are not going away, not yet, but organizations of all types, from public schools to corporations, are opting to train and inform via the web. (2014, para. 1)

As with any market, increases in interest and demand are accompanied by a surge in products. This has proven true in the online training market. With increased interest in providing online employee training, a proliferation of online training providers and materials has emerged.

Choosing Online Training Products

While the convenience, efficiency, and affordability of online training can be attractive to organizations that need to deliver employee training (Santerre, 2005), choosing a commercially available training product that meets specific contextual training needs can be difficult. As selling points, many e-learning providers aim to produce training materials that have unique features or particular areas of focus. These distinct characteristics can make the training product more marketable, but if the unique focus or features are not needed or appropriate within a particular training setting, it can also limit the ability of that product to meet the training needs within a specific workplace environment. Thus, before investing time and money into a program, it is important that training purchasers effectively and efficiently evaluate online training products to determine if they meet the specific contextual needs of the organization and the employees (Murphy, Keiffer, Neal, & Crandall, 2013).

In addition to examining online training programs relative to contextual needs, training purchasers should also determine the instructional soundness and quality of the online training product. As in any market, the quality and costs of online training products vary. Strother (2002) offers a warning to uninformed training purchasers who may focus on cost-effectiveness over quality by restating the training adage, “Wise training managers realize the bitterness of poor quality remains long after the sweetness of low price has been forgotten” (para. 7). As this saying indicates, in addition to affordability considerations purchasers must evaluate instructional quality when choosing an online training product. Inspecting online training programs to determine adherence to known quality indicators is a key step in the purchasing process, but many decision makers lack familiarity with best practices in online training or an understanding of what constitutes quality in an online instructional environment (Murphy et al., 2013).

Need for an Evaluation Tool

Even though online learning has become more accepted by those in charge of making training decisions (Batalla-Busquets & Pacheco-Bernal, 2013), the selection of a high-quality and contextually appropriate online training program can be a daunting task. The proliferation of online training materials paired with the lack of an online training background has left many decision makers in a quandary when tasked with purchasing e-learning products (Barker, 2004, 2007; Murphy et al., 2013; Neal et al., 2011; Zaied, 2012). According to Grollman and Cannon (2003), a high-quality and contextually appropriate e-learning program can be as cost effective and efficient as face-to-face training, but how do decision makers who are untrained in the evaluation of e-learning determine which products are contextually appropriate and of high quality?

Based on the aforementioned dilemma, a tool is needed that assists decision makers in the e-learning evaluation process, allowing them to make informed training purchases. A good online training evaluation tool should make it easy for the decision maker to evaluate training programs relative to quality and to assess the program's ability to address the specific needs of the company (Neal et al., 2011). The tool should also be easy to use, include numeric data for valid objective comparisons, and take into consideration the time required of busy purchasers to complete the evaluation (Neal et al., 2011; Pisik, 1997).

The remainder of this chapter describes the creation, testing, and application of such an instrument. The Customizable Tool for Online Training Evaluation (CTOTE) is an evaluation instrument developed to help decision makers assess multiple online training programs against known best practices and consider context-specific training needs via a weighting process.

Instrument Development

As indicated above, there is a need for an online training evaluation tool that allows potential purchasers to perform multiple levels of evaluation, as e-learning is a multidimensional construct (Agariya & Singh, 2012). As its primary function, 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 (Murphy et al., 2013; Zaied, 2012). This includes evaluating areas such as use of relevant media, intuitive interface design, and appropriate assessment methods. The evaluation tool must also help the decision maker determine if the training program under consideration addresses the overarching training needs

(Murphy et al., 2013; Strother, 2002). This involves an examination of the training outcomes, content, and activities to ensure they are appropriate and will help the organization meet regulations and achieve goals. 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 and employees (Becker, Fleming, & Keijsers, 2012; Istrate, 2013). This involves consideration of what aspects are most important within specific organizational contexts, as different work sites and employees possess unique training needs (Murphy et al., 2013).

The development of the Customizable Tool for Online Training Evaluation (CTOTE), an instrument that can perform the aforementioned tasks, transpires using a three-step process. During the first step, the authors slightly modify and test an existing instrument to determine the effectiveness of the instrument in performing the three key tasks. Results of this action are used in the second step to inform a complete overhaul of the modified instrument, which results in the current CTOTE tool. This new instrument is then tested to ensure content validity. The third step in the development of the instrument involves a new and unique feature, the inclusion of a Delphi panel to inform a customization process. A weighting procedure is devised that incorporates data derived from the panel to form the customization element of the instrument (Murphy et al., 2013). Each of these processes are described in the sections below.

Modification of Established Instrument and Testing

While working for AT&T as a performance technologist in 1997, Ginger Pisik recognizes the importance of evaluating the instructional soundness of online training. Regardless of whether an organization is considering the purchase of retail online training packages or taking on the task of developing online courses in-house, she asserts that diligent and consistent evaluation must be performed. However, she also acknowledges that the person in charge of making training decisions and purchases (managers) often does not have the time nor the skills to perform thorough evaluations.

To facilitate efficient and effective evaluations of online training materials Pisik creates the AT&T Online Evaluation Form. This tool uses a ranking method to allow managers to fairly and consistently evaluate online training courses. The form is a 68-item instrument that ultimately assesses five critical areas for consideration when choosing an online training program: content and instruction, learners, job transfer, design and packaging, and operation (Pisik, 1997). Numeric scores are calculated for each of these five critical area sections and are then combined to formulate an overall rating for each online training program evaluated. The overall ratings can then be used by decision makers when considering multiple programs for purchase.

Modification of the Pisik Instrument

The original Pisik (1997) instrument provides untrained decision makers a tool that facilitates efficient evaluation and comparison of multiple online training programs.

Using this tool, managers can determine strengths, weaknesses, and usability of an online training program, as well as compare findings across multiple programs. For this reason, the authors of this chapter chose to test the ability of this existing instrument to help decision makers assess the three key areas for consideration during the purchase of commercial online training that were mentioned earlier which include quality, ability to meet training needs, and the ability to meet contextual needs.

However, in order to use the older instrument modifications were necessary. Technologies have changed since 1997, and a few items within the instrument needed to reflect these changes. As such, wording within the instrument was revised to delete references to obsolete technologies (Neal et al., 2011). Additionally, five demographic questions were added to the instrument to assist the authors in understanding the ability of multiple demographic groups to use the tool effectively. The result was an updated Pisik (1997) instrument that contained five general demographic questions in the first section, followed by the original 68 items, some of which had been slightly reworded to remove references to obsolete technologies.

As indicated above, the instrument contained five areas of emphasis: content and instruction, learners, job transfer, design and packaging, and operation. These five areas included 68 statements used to judge an online training program. The first and largest area, content and instruction, contained a series of statements related to areas including but not limited to the objectives, content, instructional methods, strategies, and available help found within the online program. The second area contained a series of statements that examined learner activities within the training module, including the amount of learner control and engagement. The third area of questions concerned the ability of the information provided within the training program to be easily used and transferred to the user's work environment. Design and packaging (the fourth area) consisted of statements concerning the layout, grammar, clarity, and organization of the training material. Lastly, the fifth area looked at the operation of the program such as ease of navigation and clarity of instructions (Neal et al., 2011; Pisik, 1997).

Testing the Modified Instrument

To assess the ability of the updated instrument to assist decision makers in the evaluation of online training, a suitable context had to be identified. As previously mentioned, many decision makers need assistance when determining which existing online training to adopt (Murphy et al., 2013). This is particularly true within the food service industry. While multiple online food safety training programs are available, the challenge for food service decision makers is evaluating programs to determine which, if any, are instructionally sound and meet the needs of the organization (Egan et al., 2007; Murphy et al., 2013). These factors provided the authors with the perfect setting in which to test the effectiveness of the revised Pisik (1997) instrument.

To test the updated instrument within a food service environment, 37 subjects who were well versed in hotel and restaurant management completed three different food safety online training modules. The modules were part of the interactive food safety training program developed by Environmental Health Australia and covered topics that are commonly required of food safety training including health, hygiene, and temperature measuring (Neal et al., 2011). The participants were then asked to use the updated Pisik (1997) instrument to evaluate each module and complete the study by answering open-response questions that indicated the degree to which they were able to assess and evaluate the online food safety modules with the modified Pisik (1997) instrument.

Results indicated that the updated instrument did allow for numeric comparisons on best practice aspects of the three online training modules. As examples, items within the content and instruction section of the instrument received positive responses, with 90 % of participants ranking all three modules "excellent" or "very good" in this area. Conversely, only 67 % rated items from the job transfer section of the instrument positively (Neal et al., 2011). This demonstrated that the instrument helped decision makers consider the ability of the online modules to meet the content training needs of food service workers. Thus, participants were able to broadly rank the strengths and weaknesses of the three online food safety training modules relative to quality and appropriateness of content using the updated instrument.

However, open-response data indicated respondents felt the instrument was not thorough enough to provide a consistent assessment of the online training, which was also evidenced in instrument response data with inconsistent overall ratings amongst participants. Respondents also suggested the instrument was not detailed enough to factor into the overall rating the specific contextual needs found in food service settings. For example, the use of realistic practice exercises (item #21) was identified as being of great importance in food safety training, while the use of various information approaches (item #14) was important, but not at the same level. However, when overall ratings were calculated, no distinctions were made and responses to both of these items were given equal weight despite having differing levels of importance in this particular training environment. This feedback suggested that the updated evaluation tool was neither strong enough nor detailed enough to reduce variability among users or address contextually specific training needs (Neal et al., 2011).

An additional finding from the open response data was that, despite wanting a more thorough instrument, participants felt the instrument should be less time consuming to complete. Although data taken from the survey indicated that the mean participant time to fill out the updated evaluation instrument was a relatively short 15 min, these experienced hotel and restaurant management participants considered this to be too lengthy and expressed the desire for a shorter tool. Furthermore, several respondents indicated they experienced confusion relative to some of the items on the instrument due to unclear wording or unfamiliar terminology.

In summary, results of this study testing the effectiveness of the updated Pisik (1997) instrument demonstrated the need for a new evaluation instrument. This new instrument should continue to be based on general best practices of online training (such as the Pisik instrument), but should also be consistent across evaluators, less time consuming to complete, clearly worded with familiar terms, and 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 (Murphy et al., 2013).

Creation of a New Instrument

To address shortcomings identified within the updated Pisik (1997) tool (Neal et al., 2011), extensive revisions that eventually resulted in the creation of a new instrument, the Customizable Tool for Online Training Evaluation (CTOTE), occurred. Because the results of the previous study indicated the updated instrument could facilitate the broad determination of the quality and appropriateness of content within online training modules, the authors initiated the development process using the updated Pisik (1997) tool as a base. The most recent instructional design wisdoms for online learning were incorporated to create new construct categories, redundancies within question items were identified and eliminated, and language, and wording were altered to be more common and reflective of online *training* as opposed to general online learning (Murphy et al., 2013). Specifics concerning these processes are provided in the following paragraphs.

Determination of Categories and Items

To determine the categories and items that should be included within the new instrument, critical design components for online learning had to be identified. To assist in identifying these components, multiple existing online evaluation tools were reviewed. The tools used in this review included a new version of the original Pisik instrument (Pisik, 2004), Quality Matters Rubric Standards (QM; MarylandOnline, 2011), Quality Standards for Evaluating Multimedia and Online Training (Gillis, 2000), and the Quality Online Course Initiative (QOCI; Illinois Online Network, 2010).

Both the Gillis (2000) and newer Pisik (2004) instruments were specifically designed to allow managers, trainers, or purchasers within business and industry to select the most effective "off-the-shelf" training software for employees. The Quality Standards for Evaluating Multimedia and Online Training instrument consisted of four essential quality categories (organizational needs, content, usability, and instructional design) and contained 80 total items. Similarly, the newer Pisik instrument included eight evaluation categories (resource requirements, registration and security, content and instruction, learners, job transfer, design and packaging, operation, and financial) and a total of 91 total items (Gillis, 2000; Pisik, 2004). While these instruments were created to assist the person in charge with choosing the most appropriate training materials and take into consideration evaluation aspects unique to workplace training, both instruments referenced outdated technologies and did not account for newer understandings and best practices as they relate to the rapidly evolving area of online learning.

To ensure that the most recent best practices for online learning were incorporated during the instrument construction process, newer instruments that were developed based on extensive research reviews of best practices in online teaching and learning were also included within the instrument review process. The 2011–2013 version of the QM rubric, which consisted of eight quality standards categories (course overview and introduction, learning objectives, assessment and measurement, instructional materials, learner interaction and engagement, course technology, learner support, and accessibility) and contained 41 weighted items was examined (MarylandOnline, 2011). Similarly, the QOCI tool which was comprised of six key categories (instructional design, communication interaction, and collaboration, student evaluation and assessment, learner support and resources, web design, and course evaluation) and included 82 individual questions was analyzed (Illinois Online Network, 2010).

The aforementioned existing evaluation tools were examined and key constructs were identified for all four instruments, and questions within these constructs were isolated. Because the revised Pisik (1997) used in the first research study served as the base for the new instrument, key constructs and items within this instrument were also determined. Once all instruments were analyzed, key constructs were examined and four key construct categories emerged (information and outcomes, content and structure, assessment and transfer, technology design, and operation). These categories were derived from the primary constructs that spanned all of the reviewed instruments (Murphy et al., 2013).

Once the four categories were determined, all questions related to the constructs from each of the instruments were added to the new instrument under the appropriate construct heading. After these items were incorporated, the all-inclusive list of questions for each of the four constructs underwent a scrutinous review.

During the review of items within each construct, redundancies within multiple questions were removed and items were consolidated to streamline the instrument. Concurrently, questions that reflected outdated practices or technologies were also removed. Lastly, wording for some items was altered to more accurately reflect the key constructs and to be more reflective of current training language and terminology (Murphy et al., 2013). Additional analyses described in the next section were performed to fine-tune the instrument, and the result was the creation of the Customizable Tool for Online Training Evaluation (CTOTE) which includes four categories and 48 survey items. The next task was to determine the content validity and reliability of the new CTOTE tool.

Determination of Validity and Reliability

To determine the content validity of the categories and items, the CTOTE instrument was tested using the method prescribed by the Indexes of Item-Objective Congruence for Multidimensional Items (Turner & Carlson, 2003), also known as the IIOC. According to Murphy et al. (2013), "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" (p. 256). The content experts that were chosen to assess the validity of the CTOTE items included six instructional designers with expertise in online training. Each of the six professionals assessed all items from the CTOTE instrument by rating each item on each objective (i.e., category) using provided operational definitions. The professionals also provided feedback on any wording or clarity issues they had with the items.

A unidimensional IIOC was conducted using the originally developed categories. The unidimensional approach assumes that each item belongs to only one category. Results of this process were analyzed to determine if any of the items fell below 0.67, a value chosen to represent agreement between four of the six instructional designers who completed the questionnaire. Based on this criterion, 13 items were identified as being of interest. The items of interest from the original categories were revised, and using the data from the original assessment the categories were rearranged to form the four categories currently available in the CTOTE (content and outcomes, assessment and transfer, technology design and requirements, and operations and support).

Subsequently, a multidimensional IIOC was conducted using the same expert panel to review the new items, categories, and operational definitions. The multidimensional approach allows for the possibility that each item may belong to multiple categories (Turner & Carlson, 2003). During this second round of testing, all of the IIOC values increased when using the multidimensional approach, with no items falling below the 0.67 agreement criterion and none belonging to more than one category. These results indicated the item and category revisions performed after the unidimensional analysis were appropriate. Thus, the IIOC results and expert judgment provided by the Delphi panel and instrument creators supported the updated items, categories, and operational definitions used in the final version of the CTOTE.

In addition to establishing content validity, a reliability study was performed to ensure the CTOTE instrument provided consistent evaluation data. The design of the study was based on a randomized crossover design. A sample of 120 reviewers (hotel and restaurant management students from a regional university) used the CTOTE to investigate two learning modules across four food safety training programs. Participants were randomly assigned to compare two food safety training programs on a module the programs had in common. For example, one participant reviewed the Hand Washing module in Program A and the Hand Washing module in Program B, while another participant reviewed the Hand Washing module in Program A and the Hand Washing module in Program C, and so on. Had all participants evaluated two modules, there would have been a total of 240 observations. However, due to incomplete response data, this study ultimately generated 195 complete responses that were used for analyses.

A principle components exploratory factor analysis was performed and supported a four category instrument structure with some items cross-loading on multiple categories. The correlations between the categories ranged from 0.38 to 0.68. Given the nature of the instrument, these correlations were in an acceptable range and supported the idea of a multidimensional instrument. Cronbach Alpha scores of 0.83 and above were calculated for all categories (see Table 1), exceeding the standard acceptable reliability coefficient of 0.7 (Nunnaly, 1978). These results

Table 1 Reliability calculation results for the CTOTE instrument

Category	*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

Notes:
Total responses for the reliability study was 195. Only complete data were used when calculating descriptive statistics within each category
Reprinted from Murphy et al. (2013), p. 257. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted with slight modification

Table 2 A comparison of categories and items across evaluation instruments

Instrument	Categories	Number of items
Original Pisik	Content and instruction	38
(Pisik, 1997)	Learners	6
	Job transfer	5
	Design and packaging	8
	Operation	11
	Total items:	68
Modified Pisik	Content and instruction	38
(Neal et al., 2011)	Learner interaction	6
	Transferability of knowledge	5
	Design and packaging	8
	Operation	11
	Total items:	68
CTOTE	Content and outcomes	17
(Murphy et al., 2013)	Assessment and transfer	7
	Technology design and requirements	12
	Operations and support	9
	Total items:	45

Note:
Reprinted from Murphy et al. (2013), p. 256. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted without modification

indicated the CTOTE instrument was able to provide consistent evaluation responses.

The results of the validity and reliability testing were encouraging and concluded the second step in the process of creating an evaluation instrument grounded in best practices of online training. Additionally, the instrument categories that emerged relative to content, outcomes, assessment, and transfer provided evaluators the power to assess which online training programs would best meet their training needs. To demonstrate the evolution of this new tool, a comparison of the categories and number of items amongst the original Pisik instrument (1997), the modified Pisik (Neal et al., 2011), and the new Customizable Tool for Online Training Evaluation (CTOTE) can be seen in Table 2. Additionally, the full CTOTE instrument including all categories and questions may be viewed in the appendix of the Murphy et al. (2013) publication.

With the core instrument sufficiently tested, the third and final hurdle was to create within the CTOTE tool the ability to assess online training in relation to meeting contextual needs within the workplace. To accomplish this task, a customization process was developed.

The Customization Process

To incorporate a method whereby the CTOTE instrument was customized so it evaluated existing online training in relation to specific contextual needs, a method was devised whereby items within the instrument could be weighted based on contextual importance. To determine contextual importance, input would be sought from an expert panel who was familiar with the specific environment in which the online training would be delivered. For example, if the online training would be used in the law enforcement industry, experts in law enforcement training would be sought for the panel. To use this process across different contexts, the panel of experts would change depending on the context for which the training was being examined. The number of experts within a panel could vary, but should number no less than 5 and no more than 15 to allow the full weighting process to unfold in a manageable fashion (Murphy et al., 2013).

Once appropriate members for an expert panel have been identified, these panel members should individually rank each item on the CTOTE instrument as most important, somewhat important, and mildly important when considering the purchase of online training to meet the particular needs within the specified context. An iterative Delphi process would then be used with the panel to compare initial results and allow for revisions to rankings until consensus was garnered across the expert panel regarding the importance level (most important, somewhat important, or mildly important) of each item. Once consensus was reached, the results of this process would be used to directly inform the weighting of items within the CTOTE instrument, with more weight given to the items deemed by the expert panel as most important (weight of 3) and less weight given to those deemed somewhat important (weight of 2) and mildly important (weight of 1).

By adding the Delphi process with an expert panel to the CTOTE tool, the result was the creation of an evaluation instrument that meets the three key needs of online training decision makers by providing an overall evaluation tool that is based on best practices of online training, considers the appropriateness of the content, and addresses the needs within a specific context as determined by experts from within that context. In essence, the expert-driven weighting process adds to the base

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)] x100 = **75%**	

Fig. 1 Unweighted example of CTOTE instrument scoring (Reprinted from Murphy et al. (2013), p. 258. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted without modification)

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/(Totals Possible overall – Total NA)] x 100							[23/(40 – 12)] x 100 = **82%**	

Fig. 2 Weighted example of CTOTE instrument scoring (Reprinted from Murphy et al. (2013), p. 259. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted without modification)

CTOTE instrument by providing a tool that can be tailored to the specific needs within a particular context and offers numeric data that can be used in cross-program comparisons. The contextualized weighting process and how it can be used to assist training purchasers in the final decision-making process are detailed in the next section.

The Contextualized Weighting Process

As an illustration of the power of the contextualized weighting process and its influence on the scoring process, Murphy et al. (2013) provided Figs. 1 and 2 which demonstrated how scores within a category were impacted once the weighting system was applied. In this fictitious example, no weighting was incorporated in the CTOTE ratings that appear in Fig. 1, and the instrument was scored without the benefit of the contextualized Delphi input. The red circles represented the rating that was given for items one through five by the fictitious evaluator. The end result based on unweighted calculations was an overall category score of 75 %.

However, when Murphy et al. (2013) applied weighting to the same fictitious example in Fig. 2, the overall category score raised substantially. In addition to the same ratings that appeared in Fig. 1, each item in Fig. 2 was also assigned a weight based on results of a fictitious Delphi panel. The lowest weighted items were three and five (1), while items one and two were given the highest weight (3) and item four (2) fell in between. As a result of the weightings, the overall category percentage changed significantly. More specifically, items two and five were ranked the same (4) in both examples, but when weights were added (Fig. 2), item two became much more important to the overall category total. This led to a raise in the overall category percentage, increasing from the unweighted example of 75 % (Fig. 1) to a weighted 82 % (Fig. 2). Specifics on instrument calculations that occurred to compute weighted category percentages are presented in the next section.

Instrument Calculations

As noted in the bottom portion of Fig. 2, to calculate the weighted percentage for a category you must determine the total points accumulated in that category, but you must also compute the total possible points as well as take into account the total number of not applicable answers (Total NA) that were provided. To determine the total points, multiply the reviewer ranking (in this case 0–4) by the weight (1–3) for each item, then add the results to get the total points for the category (Murphy et al., 2013).

To determine the total possible points for each item, multiply the maximum ranking possible (in this case 4) by the weight (1–3). The total possible points for the category are derived by adding the results for each item within the category. Additionally, to prevent questions that were deemed Not Applicable from influencing the overall category score, subtract the total NA scores for all items from the total possible. Total NA scores are equal to the total possible score and are calculated only for items that are rated as NA by the evaluator (Murphy et al., 2013).

The final step in calculating the overall category percentage is to divide the total points received from the total possible (minus the NAs) and multiply by 100. The resulting formula is [Total/(Total Possible − Total NA)] × 100. This calculation provides a weighted percentage for this specific category that can be compared to the same weighted percentage for this category when the evaluator reviews additional online training programs. In addition to comparing category percentages, evaluators can also calculate and compare the overall rating scores for multiple programs. To do this, the evaluator must compute the overall percentage score, which is an average of the four category percentage scores. Because weighting is performed at the category level, additional weighting is neither needed nor appropriate in the calculation of the overall score. Rather, to get the overall percentage score, simply add the category percentages for all four categories together and divide by four. The resulting formula is [(Category 1 Percentage + Category 2 Percentage + Category 3 Percentage +

Category 4 Percentage)/4]. As with the category percentages, this overall percentage score can be used to compare multiple online training programs (Murphy et al., 2013).

As the example presented in Figs. 1 and 2 indicated, the weighting system offered the ability to "customize" the instrument to focus on and emphasize the importance of specific organizational needs. Through this customization process the decision maker was able to obtain more precise and applicable information when evaluating multiple online training programs. While the aforementioned fictitious example was

Operations and Support Category	SA	A	DA	SD	N/A	Weigth	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 trainess' 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. 3 Expert Delphi weightings in operations and support category (Reprinted from Murphy et al. (2013), p. 268. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted with modifications)

informative, the ability of real-world experts to participate in the Delphi weighting method had to be tested.

Testing the Delphi Weighting Process

A test of the Delphi weighting process was performed to verify if the process could be used by experts within the field. For this specific study, as with the previous instrument-related research, the authors focused on the food service industry. In particular, food safety online training served as the specific context. The expert panel utilized to test the Delphi weighting process consisted of 12 experienced retail food service managers who were familiar with food safety training (Murphy et al., 2013). The experts were provided directions to rate each item on the CTOTE in relation to its importance when considering the purchase and effectiveness of food safety training in retail food service. After only two Delphi iterations, the panel was in consensus on the ratings of each item.

To demonstrate the outcome, Murphy et al. (2013) provided the resulting actual weightings from this panel within the Operations and Support category of the CTOTE instrument. These weightings can be viewed in Fig. 3. As illustrated in this example, the expert Delphi group was able to effectively use the Delphi weighting process. Via the Delphi ratings, the panel indicated that the provision of clear directions and the ability to print important information were of greater importance when choosing an online training food service program than other items assessed within the Operations and Support category. As a result, when an overall rating score is tallied within this category, these higher-ranked items will be given more consideration and weight. An in-depth description of the calculation process used to obtain the overall rating score is provided in the next section.

With an understanding of how to compute weighted scores using the CTOTE tool and verification that the Delphi weighting system can be utilized in the field, the last task was to test the effectiveness of this new instrument in providing a more thorough, consistent, and contextualized assessment of multiple online training programs.

Testing the CTOTE Instrument

As described in previous sections, content validity and reliability of the CTOTE tool were established, and the usability of the Delphi weighting system was verified. The next step was to test the new CTOTE in relation to an established instrument to examine the impact of the updated categories, items, and weighting procedure on ratings of actual retail online training programs. In other words, was the new instrument capable of providing more thorough and contextualized evaluations than an existing instrument? The final step involved testing of the CTOTE across multiple training programs to ensure the new tool could provide consistent evaluation data. Once again, the authors chose the food service industry and, in particular,

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[a]	14.38	
Pisik, 1997	22	87.12[a]	11.14	
Correlations				
CTOTE – no weight vs. weighted	25			$r = .90$[a]
CTOTE – no weight vs. Pisik	20			$r = .55$[a]
CTOTE – weighted vs. Pisik	21			$r = .47$[a]

Notes:
Reprinted from Murphy et al. (2013), p. 260. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted without modification
[a]Statistically significant at alpha = .05

food safety online training as the context for this round of testing. Descriptions and results of these tests are provided in the sections below.

Testing Across Multiple Instruments

Testing across instruments was performed by comparing evaluation results derived from the new CTOTE tool to results obtained through the use of the original Pisik (1997) instrument. Additionally, weighted versus unweighted scores on the CTOTE were compared. These comparisons allowed the authors to determine if the item and category changes made on the CTOTE influenced overall rating scores as compared to an established instrument, as well as ascertain the impact on ratings within the CTOTE scores that resulted from the weighting procedure (Murphy et al., 2013).

To test results from the CTOTE instrument in relation to results from the Pisik (1997) instrument, 26 reviewers examined a retail online food safety training program. The reviewers then used both the Pisik and CTOTE instruments to evaluate the online food safety training program. Scores, means, and standard deviations were calculated for each instrument. Because data in some areas were not recorded, calculations only occurred in instances where all values were present. Using matched data, correlations were then calculated. This analysis determined if the results from the CTOTE instrument (weighted and unweighted) were reporting the same conclusions as those reported by the established Pisik instrument (Murphy et al., 2013).

Correlations within the CTOTE tool between the weighted and unweighted versions were very strong and positive, $r(23) = .90$, $p < .05$. Correlations were not as strong between the unweighted CTOTE instrument and the Pisik instrument, but were also positive and statistically significant $r(18) = .55$, $p < .05$. Lastly, correlations were still positive between the weighted CTOTE instrument versus the Pisik, but this comparison had the lowest significant correlation at $r(19) = .47$, $p < .05$. The findings demonstrated that the CTOTE (weighted and unweighted) and

Table 4 Breakdown of reviewer assignments

	Retail online food safety training programs				
Review groups	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*

Note:
Reprinted from Murphy et al. (2013), p. 261. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted without modification

Pisik (1997) instruments were significantly correlated, but the difference between the weighted CTOTE and the Pisik was greater than the unweighted CTOTE when correlated with the Pisik instrument (Murphy et al., 2013). Results of the analyses including means, standard deviations, and correlation coefficients may be viewed in Table 3.

Consistency Testing Across Multiple Programs

To test how consistently the CTOTE instrument would perform when examining multiple online training programs, 48 reviewers examined four similar retail online food safety training programs. The programs covered the same general content, and each reviewer compared two of the programs. Weightings that were derived from the expert food service Delphi panel during earlier testing of the weighting process were incorporated, as they were contextually appropriate for this study. Reviewers used the CTOTE weighted instrument to evaluate both programs, producing a total of 96 CTOTE reviews (Murphy et al., 2013). Table 4 contains a breakdown of reviewer assignments used during this testing process.

Using the calculation processes described in an earlier section, overall percentage scores were computed for each program review that was performed. Overall program percentages were then compared across all four programs. Based on these overall percentage comparisons, it was easy to identify which online training programs the reviewers preferred. The resulting preferences can be seen in Table 5. Data indicated the programs were ranked consistently by raters, with Program 3 or Program 4 consistently ranked lower than Program 1 and Program 2. Program 1 was the highest ranking program, barely edging out Program 2, while Program 4 was clearly the lowest ranking program (Murphy et al., 2013).

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[a]	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

Notes:
Reprinted from Murphy et al. (2013), p. 262. Available at http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242 under Creative Commons Attribution 3.0. Reprinted without modification
[a]Two reviewers mistakenly reviewed Program 1 vs. 2 instead of 2 vs. 4 and 3 vs. 4

Discussion of Testing Results

Results of both tests indicated that the weighting process used by the CTOTE was effective in making strengths and weaknesses of online training programs more noticeable. Findings also indicated the CTOTE tool was able to provide consistent ratings across multiple programs and reviewers. Both of these outcomes are briefly explored below.

The Weighting Impact

As previously indicated, results between the weighted and unweighted versions of the CTOTE instrument were strong and positive. Results between the Pisik (1997) instrument and the CTOTE instrument (weighted and unweighted) were not as strong, but were also positive and statistically significant (Murphy et al., 2013). Findings demonstrated when the same program is rated by these instruments, the overall scores were highest for the Pisik (1997) instrument and lowest for the weighted CTOTE instrument (Murphy et al., 2013). This statistically significant difference demonstrated that when weighting was applied via the new CTOTE tool, the ratings were more extreme (higher or lower) because a wider range was used. In other words, if a program excelled in important areas it was easier to get a high score, but it was more difficult to get a high score if a program fell short in an important area. In both situations the weighting process emphasized strengths or weakness in important areas, more so than in unweighted ratings. This emphasis resulted in a noticeable and significant difference in the overall percentage score of the reviewed program. These findings support the use of weighting to facilitate a more specific contextual evaluation, thus providing the reviewer more contextually relevant data on which to make purchasing decisions relative to online training (Murphy et al.).

Consistency in Ratings

Results of the testing also indicated multiple food safety programs were consistently ranked by numerous raters, demonstrating the ability of the CTOTE instrument to provide consistent evaluation ratings. Similarly, the use of the expert Delphi panel to inform weighting was effective in accentuating small but potentially critical contextual differences between four very similar online food safety training programs. These differences were magnified by this weighting process, making them more noticeable in the overall percentage scores.

The findings from the CTOTE instrument testing indicated that this new tool was effective at addressing the contextual and consistency concerns that were identified when Neal et al. (2011) tested the revised Pisik (1997) instrument. While these results were promising, a final round of testing with a larger industry audience was desired to solidify the CTOTE tool as a strong and consistent evaluation instrument.

Food Service Industry Testing

As previously discussed, the food service industry necessitates the provision of food safety training by employers. Improper food handling by food service employees is a growing issue (Almanza & Nesmith, 2004), but training employees who come from such diverse educational and ethnic backgrounds is a challenge (Neal, Dawson, & Madera, 2011). Similarly, classroom facilities are typically not available in food service locations to facilitate face-to-face training, making the delivery of training difficult. Both of these challenges make the use of online food safety training programs an attractive option to food service managers, as online programs can offer training in multiple languages and can be delivered anywhere an internet connection exists (Howton et al., 2016). Unfortunately, it is difficult for managers to determine which training program will meet their particular needs (Frash, Binkley, Nelson, & Almanza, 2006). Given the aforementioned circumstances, it was appropriate to perform industry testing of the CTOTE tool within this specific environment.

The main objective of the Howton et al. (2016) study was to assess and compare four commercially available online food safety training programs commonly used in the food service industry and determine which program was preferred among front-line food service employees. The CTOTE instrument was used to facilitate the comparison of the four training programs, and weightings derived from an expert Delphi food service panel in a previous study (Murphy et al., 2013) were used to provide the appropriate instrument weightings for this context.

The CTOTE evaluation tool was provided to 96 participants who were representative of front-line food service employees, with 87 choosing to take part in this evaluation study. A random sampling method was used to assign each participant to review two of the four possible online food safety programs (Program A, B, C, & D) in a similar manner as used by Murphy et al. (2013) to test for instrument consistency. Access to the assigned programs was provided based on the random assignments, and participants were asked to use and fully evaluate the first program (in the order of their choosing) using the CTOTE tool prior to accessing the second program. After completing the first program evaluation, participants were asked to evaluate the second program using the same procedure. All 87 individuals successfully completed evaluations for both assigned programs.

Calculations of the instrument ratings were performed as described in a previous section, and the resulting percentage data enabled the ranking of the four programs as a whole, and by CTOTE category. Program A received the highest overall rating from respondents, followed by B, D, and then C, indicating that Program A was preferred amongst study participants. Within the four categories, the percentage ratings were reflective of the overall ranking, with Program A ranked first in two categories (content and outcomes and operations and support) and second in the remaining two categories (assessment and transfer and technology design). As would be expected by the overall ratings, Program C had the lowest category rankings with four categories ranked fourth and one category ranked third (Howton et al., 2016).

The findings of this industry-based study revealed that the CTOTE tool was effective in allowing participants to rate four commonly used commercial online food safety training programs. The weighting procedure allowed for differentiation among these similar programs such that a clear ranking was obtained, and the successful completion of two program evaluations by all 87 participants illustrated that the instrument was easy to use. Thus, the results suggested that the CTOTE tool was used effectively within the food service industry to evaluate commercial online food safety training programs.

Summary

With projections that the online learning market to will garner $107 billion in 2015 (McCue, 2014), and the increased acceptance of online training within business and industry (Batalla-Busquets & Pacheco-Bernal, 2013), the purchase and use of commercially available online training is following an upward trajectory. Accompanying this increase is the need to support uninformed training decision makers so they may effectively and efficiently assess multiple online training programs.

When considering commercial online training programs for purchase, decision makers should evaluate the instructional soundness of the online training products, as the quality of retail online training products can vary. Determining adherence to known quality indicators is a key step in this evaluation process, but many decision makers may not understand what constitutes quality in an online instructional environment (Murphy et al., 2013). A tool that easily assists these decision makers in evaluating online training programs against known best practices for online learning would be beneficial.

In addition to evaluating the quality of an online training program, potential purchasers must also consider the contextual appropriateness of a commercial program. Elliott Masie succinctly describes 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) assert specified contexts must be considered when considering quality in e-learning, but Mosharraf and Taghiyareh (2013) indicate that vendors are unable to tailor training to meet the needs of all contexts. Therefore, in order to make appropriate online training purchasing decisions, persons in charge must be able to determine if existing online training programs meet the contextual needs of their specific business or industry (Murphy et al., 2013).

As indicated by researchers (Barker, 2004, 2007; Seufert, 2002; Zaied, 2012), it is difficult for companies to decide which e-learning product to choose (Murphy et al., 2013). To assist in this process, the authors sought to develop and test an instrument that would support uninformed decision makers in the evaluation of multiple retail online training programs. The result was the creation and testing of the Customizable Tool for Online Training Evaluation (CTOTE), an instrument that takes 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 et al., 2012; Istrate, 2013).

Over a span of three separate research studies, the CTOTE instrument was developed, validated, and tested within the area of food service, an industry where decision makers may be ill-equipped to choose the best training options (Egan et al., 2007; Neal et al., 2011). When multiple online food safety training programs were evaluated, the CTOTE instrument produced consistent ratings across raters, which demonstrated the consistency of the tool. However, because the weighted CTOTE instrument generated a larger percentage score range, the weighted CTOTE ratings magnified even the smallest contextual differences. Similarly, when CTOTE ratings were compared to those of an established instrument, results indicated the instruments rated the online program similarly, but the CTOTE instrument allowed for more discerning distinctions.

Strengths and Limitations

A strong point of this new instrument is that it is firmly based upon researched best practices, but perhaps more significant is that the CTOTE has withstood stringent initial testing. Two rounds of rigorous IIOC testing occurred to substantiate the content validity of the categories and items within the CTOTE, and a large randomized crossover design study was used to firmly establish reliability. The instrument was also substantiated as a user-friendly tool when it was effectively tested within a specific industry setting. Additionally and potentially the largest strength of the CTOTE is the novel use of industry-specific Delphi groups to facilitate contextually appropriate weighting of evaluation data. This innovative feature distinguishes the CTOTE from previous tools and affords decision makers the ability to account for context-specific needs and factors.

Despite the aforementioned strengths, there are potential limitations that could impede the use of the CTOTE instrument. For example, the only way the tool can be effective is if decision makers take the time to thoroughly examine each program they are considering for purchase. Although the CTOTE instrument can be filled out relatively quickly, decision makers must review each program in order to provide the information required by the CTOTE to calculate accurate comparison data. The reviewing task could be viewed as too time consuming by training purchasers, and the CTOTE instrument could thereby be dismissed as a viable decision-making aid. Additionally, the Delphi process used to contextually weight responses could be viewed as onerous. The person in charge must coordinate the Delphi process, but this task could be intimidating for the decision maker, which could impede the implementation of the CTOTE instrument. Thus, some of the very things that are considered strengths of the CTOTE instrument could also be viewed as obstacles to its implementation.

Conclusion

All of the findings from instrument and industry testing of the CTOTE instrument culminate to support the determination that the CTOTE tool facilitates consistent ratings of commercial online training programs that provide more discriminating data on which a decision maker can make clear and informed purchasing decisions. Currently, use of the CTOTE is expanding beyond food service to include the evaluation of online law enforcement training. It is recommended that future studies examine the use of the CTOTE instrument across multiple contexts to document that the findings from this food service research are indeed transferable.

Acknowledgments This research was supported in part by an USDA National Integrated Food Safety Initiative Grant to the authors.

References

Agariya, A. K., & Singh, D. (2012). e-Learning quality: Scale development and validation in Indian context. *Knowledge Management & E-Learning: An International Journal (KM&EL), 4*(4), 500–517. Retrieved from http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/139/163

Alkhattabi, M., Neagu, D., & Cullen, A. (2010). Information quality framework for e-learning systems. *Knowledge Management & E-Learning: An International Journal (KM&EL), 2*(4), 340–362. Retrieved from http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/21/62

Almanza, B. A., & Nesmith, M. S. (2004). Food safety certification regulations in the United States. *Journal of Environmental Health, 66*(9), 10–14.

ATD Research. (2015). *2014 state of the industry report*. Alexandria, VA: Association for Talent Development.

Barker, 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

Barker, N. K. (2007). E-learning quality standards for consumer protection and consumer confidence: A Canadian case study in e-learning quality assurance. *Journal of Educational Technology and Society, 10*(2), 109–119. Retrieved from http://www.ifets.info/journals/10_2/10.pdf

Batalla-Busquets, J. M., & Pacheco-Bernal, C. (2013). On-the-job e-learning: Workers' attitudes and perceptions. *The International Review of Research in Open and Distributed Learning, 14*(1), 40–64. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1304/2444

Becker, K., Fleming, J., & Keijsers, W. (2012). E-learning: Ageing workforce versus technology-savvy generation. *Education + Training, 54*(5), 385–400. 10.1108/00400911211244687.

Caudill, J. G., & Reeves, B. (2014). Strategic management of workplace e-learning. In S. Hai-Jew (Ed.), *Remote workforce training: Effective technologies and strategies* (pp. 17–27). Hershey, PA: IGI Global.

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. doi:10.1016/j.foodcont.2006.08.001.

Frash, R., Jr., Binkley, M., Nelson, D., & Almanza, B. (2006). Transfer of training efficacy in US food safety accreditation. *Journal of Culinary Science & Technology, 4*(2–3), 7–38.

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

Grollman, W. K., & Cannon, D. (2003). eLearning: A better chalkboard. *Financial Executive, 19*(8), 45–47.

Howton, J., Keiffer, E., Murphy, C. A., Sirsat, A. S., O'Bryan, C. A., Ricke, S. C., Neal, J. A. (2016). A comparison of online food safety training programs using the Customizable Tool for Online Training Evaluation. *Food Control, 59,* 82–87. doi:10.1016/j.foodcont.2015.04.034.

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

Istrate, O. (2013). Criteria for e-learning programmes evaluation. *Journal of Advanced Distributed Learning Technology, 1*(1), 27–35. Retrieved from http://jadlet.com/index.php/jadlet/article/view/12/18

Jan, P. T., Lu, H. P., & Chou, T. C. (2012). The adoption of e-learning: An institutional theory perspective. *Turkish Online Journal of Educational Technology-TOJET, 11*(3), 326–343. Retrieved from http://www.tojet.net/articles/v11i3/11331.pdf

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

McCue, T. J. (2014, October). Online learning industry poised for $107 billion in 2017. *Forbes.* Retrieved from http://www.forbes.com/sites/tjmccue/2014/08/27/online-learning-industry-poised-for-107-billion-in-2015/

Mosharraf, M., & Taghiyareh, F. (2013). Qualitative development of e-learning environments through a learner relationship management methodology. *Knowledge Management & E-Learning: An International Journal (KM&EL), 5*(1), 56–65. Retrieved from http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/211/171

Murphy, C. A., Keiffer, E. A., Neal, J. A., & Crandall, P. G. (2013). A customizable evaluation instrument to facilitate comparisons of existing online training programs. *Knowledge*

Management & E-Learning: An International Journal, 5(3), 251–268. Retrieved from http://www.kmel-journal.org/ojs/index.php/online-publication/article/view/242/186

Neal, J. A., Dawson, M., & Madera, J. M. (2011). Identifying food safety concerns when communication barriers exist. *Journal of Food Science Education, 10,* 36–44. doi:10.1111/j.1541-4329.2011.00128.x.

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. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1541-4329.2010.00112.x/epdf

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

Pappas, C. (2015, January). The top e-learning statistics and facts for 2015 you need to know. *eLearning Industry*. Retrieved from http://elearningindustry.com/elearning-statistics-and-facts-for-2015

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. http://c.ymcdn.com/sites/aect.site-ym.com/resource/resmgr/Publication_PDFs/ET_Contents_Pages_2000-2015a.PDF

Santerre, C. R. (2005). X-train: Teaching professionals remotely. *The Journal of Nutrition, 135*(5), 1248–1252. Retrieved from http://jn.nutrition.org/content/135/5/1248.full

Seufert, S. (2002). eLearning business models: Strategies, success factors, and best practice examples. In C. Wankel & R. DeFillippi (Eds.), Rethinking management education for the twenty-first century (pp. 109–132). Greenwich, CT: Information Age Publishing.

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

Steele, J. (2015). Comparing computer-based employee training with traditional face-to-face methods. In D. Slykhuis & G. Marks (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2015* (pp. 504–507). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE). Retrieved from http://www.editlib.org/p/150043/

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). Retrieved from http://www.irrodl.org/index.php/irrodl/article/viewArticle/83/160Ubell

Training. (2014, November/December). 2014 training industry report. *Training, 51*(8), 16–29. Retrieved from http://www.trainingmag.com/sites/default/files/magazines/2014_11/2014-Industry-Report.pdf

Turner, R. C., & Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing, 3*(2), 163–171. doi:10.1207/S15327574IJT0302_5.

Van Rooij, S. W. (2011). Instructional design and project management: Complementary or divergent? *Educational Technology Research and Development, 59*(1), 139–158.

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.

Zhang, D., & Nunamaker, J. F. (2003). Powering e-learning in the new millennium: An overview of e-learning and enabling technology. *Information Systems Frontiers, 5*(2), 207–218. doi:10.1023/A:1022609809036.

Dr. Cheryl Murphy is an Associate Professor of Educational Technology at the University of Arkansas. She has served as Co-Director of the campus-wide Teaching and Faculty Support Center as well as the Director of Distance Education for her college. Her research interests include the design, development, and evaluation of online instruction, and faculty development that promotes the effective use of instructional technologies.

Dr. 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 Associate 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.

Ms. Jessica Howton received her MS degree in hospitality management from the University of Houston in 2013, and served as a graduate research assistant from 2011 to 2013. She is currently an Assistant Store Director (ASD) for a privately owned Texas based grocery retailer in their Houston Food and Drug Division. Her oversights include food and health safety, sanitation, facility management, and operations.