

# Italian translation of the questionnaire for professional training evaluation

Chiara Fregonese, Andrea Caputo and Viviana Langher

This work illustrates the psychometric properties of the Italian version of the Questionnaire for Professional Training Evaluation, designed and validated by Grohmann and Kaufeld. This 12-item questionnaire provides for the evaluation of different training outcomes, is time-efficient, applicable to several professional contexts and shows sound psychometric properties. In order to test the Italian form, we carried out two studies. In Study 1 (N = 125), an exploratory factor analysis led to a two-factor solution accounting for short- and long-term training outcomes. In Study 2 (N = 122), a five-model comparison was performed. Although at first a two-factor solution seemed to emerge, confirmatory factor analysis found the best fit in a six, inter-correlated, first-order factors model (satisfaction, utility, knowledge, application to practice, individual organizational results and global organizational results). Relationships with learning transfer, transfer quantity, type of training, training methodologies and individual variables – gender, age and tenure – are explored. Limitations and research and practical implications are discussed.

## Introduction

Professional training is one of the most widely used organizational tools for improving employees' capability to cope with existing or new challenges, develop resilience and reduce organizational and emotional difficulties (Langher *et al.*, 2017). Several frameworks to evaluate its impacts have been created, constantly expanding the levels and the depths at which training evaluation can take place (Aguinis & Kraiger, 2009; Wang & Spitzer, 2005; Wang & Wilcox, 2006). Kirkpatrick's four-level model (Kirkpatrick, 1967; Kirkpatrick & Kirkpatrick, 2016) is one of the most widely used frameworks to measure the level at which evaluation can be carried out (reaction, learning, behavior, results). Subsequently, several other research works have offered additional conceptual

---

□ Chiara Fregonese, PhD fellow, Sapienza University of Rome, Rome, Italy. Email: chiara.fregonese@uniroma1.it. Andrea Caputo, PhD fellow, Sapienza University of Rome, Rome, Italy. Email: andrea.caputo@uniroma1.it. Viviana Langher, Associate Professor, Sapienza University of Rome, Rome, Italy. Email: viviana.langher@uniroma1.it

frameworks and tools aimed at taking into account multiple training and organizational levels of evaluation (Alvarez *et al.*, 2004; Holton, 2005; Kraiger, 2002; Phillips, 2012; Sitzmann & Weinhardt, 2017). However, at the moment, assessing the impacts of training remains a challenge for many learning and development (L&D) practitioners, who often find it hard to go beyond the simple participants' immediate reactions at the end of a training program (Nickols, 2005; Pineda, 2010) for cost, professional approach and even mindset reasons (Lewis, 1996; Phillips *et al.*, 2015; Swanson, 2007; Wang & Wilcox, 2006).

A call for valid and usable training evaluation tools has been made by many authors and institutions (Aguinis & Kraiger, 2009; ASTD, 2009; Guerci & Vinante, 2011; Phillips & Phillips, 2016). In response, numerous models and instruments have been created with the purpose of evaluating the impact of professional training.

Ambitious and comprehensive evaluation approaches and methods have been designed to evaluate training impacts through financial quantification and return on investment (ROI) estimation (Phillips, 2007; Phillips & Phillips, 2016), storytelling and qualitative inquiry (Brinkerhoff, 2006) and involved stakeholders' satisfaction (Guerci & Vinante, 2011). Although such methods represent fundamental attempts at quantification and the creation of conceptual and practical guidelines for L&D practitioners, the use of inventories and scales nevertheless remains very useful when it comes to guaranteeing efficiency and for comparison among different training interventions and environments.

Among the available quantitative tools, Holton and colleagues (Bates *et al.*, 2012; Holton, 2005; Holton *et al.*, 1997; Holton, Bates, & Ruona, 2000) pioneered the research around a scale that could identify barriers and enablers of learning transfer, from a motivational, environment, training and ability point of view. Their research work produced a generalized Learning Transfer System Inventory (LTSI), comprising a set of 112 items, which encompasses several organizational, individual and managerial dimensions. This scale focuses on the perceived transfer climate before and after the training more than on the training intervention outcome itself and thus serves the purpose of wider organizational intervention and transformation analyses.

Focusing more on the evaluation of training impacts, Ritzmann *et al.* (2014) built the Training Evaluation Inventory, which is a 45-item scale measuring subjective enjoyment, usefulness, difficulty, knowledge gain and attitude toward training, as well other training design dimensions. This scale, albeit very complete and capable of comparing different training interventions, has a length which makes it time consuming, and it is hard to embed in other more context-specific evaluation measures.

In contrast, Grohmann and Kauffeld (2013) produced and validated a concise Questionnaire for Professional Training Evaluation (Q4TE). The scale measured short- and long-term training outcomes (Kirkpatrick & Kirkpatrick, 2007; Wang & Wilcox 2006), with a six, inter-correlated factor model, including satisfaction, perceived utility and knowledge as short-term outcomes, and application to practice, organizational individual results and global results as long-term outcomes. The Q4TE has sound psychometric properties and combines shortness (12 items) with measurement of training impacts at different levels and in different terms: from reaction (global satisfaction and perceived utility) to knowledge, application to practice and organizational results (individual and global).

One of the main advantages of this scale is that it is extremely time- and cost-efficient, which makes it suitable for use in combination with other quantitative or qualitative measurement tools and within wider organizational evaluation initiatives. Indeed, it shows good applicability to different training contexts, regardless of the specific training contents. Moreover, its psychometric stability and reliability make it a useful tool for research purposes as well.

Even though the scale has been widely used in German, English and other languages (Grohmann *et al.*, 2014; Lourenço, 2016; Massenber *et al.*, 2015; Rekik & Bali, 2017), no similar scale has been validated in the Italian literature to date. This is why, based on the previously discussed advantages, we chose to focus on the Q4TE.

This paper aims at translating the Q4TE into Italian and at investigating its validity and reliability in this context for use in both organizational research and practice.

## Current research questions

Through this study we aimed at validating the Q4TE scores within the Italian context by exploring the current underlying factor structure and by examining the differential and the discriminant validity of the scale, as well as the role of individual and training-related factors as possible covariates.

### Underlying factor structure

In the exploratory phase of the original scale validation (2013), Grohmann and Kauffeld identified two factors underlying the Q4TE, which were, respectively, referring to short-term (satisfaction, perceived utility, knowledge) and long-term evaluation (application, individual organizational results and global organizational results). In their confirmatory factor analyses, however, based on modification indices, residual variances and item wording inspection, the authors developed the final version of the Q4TE, which resulted in six subscales made of two items each. The authors also investigated other factor structures: a two-second-order latent factor model (short- and long-term results), a four-factor model (based on Kirkpatrick's model of Reaction, Learning, Behavior, Results) and a six, inter-correlated factor model (Satisfaction, Utility, Knowledge, Application, Individual Organizational results and Global Organizational Results). However, only the six, inter-correlated factor model led to better goodness-of-fit indexes.

Thus, our first research question aims at discovering whether these results apply also to the Italian validation of the Q4TE scores and, if not, which model best represents the factor structure of the scale.

### Differential and discriminant validity

In line with the tools chosen by the authors of the scale, to gain evidence for the capability of the questionnaire to detect actual training impacts and learning transfer, we aimed at identifying whether the Q4TE scores were associated with concrete learning transfer and, also, with learning transfer quantity.

For this reason, our second research question concerns the differences between respondents who report to be successful in transferring the learning into practice and respondents who do not. We expect the former to show higher Q4TE scores than the latter.

Our third research question aims at exploring the relationship between the Q4TE scores and transfer quantity, that is, the number of times training content is applied at work (Grohmann & Kauffeld, 2013; Kauffeld & Lehmann-Willenbrock, 2010; Kauffeld *et al.*, 2008). Consistently with the original validation study, we expected the questionnaire to be associated with a higher incidence of training content being applied at work.

Moreover, our fourth research question aims at identifying the influence of other possible variables such as gender, age, organizational tenure, type of training (i.e. open skills, closed skills or both; Blume *et al.*, 2010; Yelon & Ford, 1999), training course duration and also the variety of training methodologies used during the training (i.e. face-to-face lectures, experiential exercises, group and peer discussion, simulations, virtual learning sessions).

In order to answer these questions we carried out two studies, both with adult working populations. Study 1 ( $N = 125$ ) aimed at answering to our first research question, whereas Study 2 ( $N = 122$ ) answered to our second, third and fourth research questions. To test construct validity, we used Study 1 for exploratory factor analysis (EFA) in order to examine the factor structure of the scale; Study 2 was then used for confirmatory factor analysis (CFA) to verify the factor structures found in Study 1.

# Methodology of Study 1

## Translation of Q4TE into Italian

For the translation of the Q4TE into Italian, we used a four-step methodology (Sousa & Rojjanasrirat, 2011) starting from the English translation provided by the authors (Grohmann & Kauffeld, 2013):

- Forward translation: Three linguistics experts independently translated the original scale from English to Italian.
- Reconciliation: The three experts and researchers compared the forward versions with the original scale and reconciled their differences.
- Back translation: A linguistics expert with a good proficiency in both Italian and English translated the reconciled version from Italian back to English.
- Comparison: Finally, the original and back-translated versions were newly examined. The back translation was compared with the original scale to determine whether or not any difference between the English and Italian versions existed in meaning and concept coherence. After the necessary corrections, the tool reached its final Italian language version. Through this approach, concept and language equivalence were guaranteed.

### Content validity

After the translation process was completed, some experts assessed the content validity of the Italian version. Three psychologists, experts in organizational training, were asked to evaluate the item content, meaning and clarity by assessing the proposed items as *inappropriate*; *appropriate to some extent-item should be revised*; *appropriate-minor changes required* and *quite appropriate* (McKenzie *et al.*, 1999). We considered the assessments *appropriate-minor changes required* and *quite appropriate* as acceptable, and content validity ratio (CVR) was calculated accordingly (Lawshe, 1975).

### Face validity

To test the face validity and comprehensibility of the tool, Q4TE was pre-tested with 10 volunteer participants who had attended a professional training in the last 2 years. These volunteers assessed Q4TE items for readability, comprehensibility, sentence length, clearness and clarity of meaning. After pre-testing, no change to the content of the Italian version was required. The final version is shown in Table 1.

### Participants

A convenience sample of 125 participants was recruited (75 women and 50 men) whose mean age was 38.03 (SD = 9.01). We considered only participants who had attended a professional training course that dated back between 4 weeks and 2 years, in order to allow knowledge to be turned into practice and still keep good learning content retention (Grohmann & Kauffeld, 2013; Linton, 1982). We partnered with an Italian training provider in order to collect a wide sample in a short time. A web-based survey was promoted at least 4 weeks after each training course had finished. The survey was conducted according to the online survey design, development and implementation guidelines suggested by Andrews *et al.* (2003) and was chosen because of its ease of administration to geographically disperse respondents across the national territory (Evans & Mathur, 2005).

The sample included employees from several industries (i.e. energy, pharmaceutical, telecommunication, transports, finance) mainly referring to private sector. For the present study 100 per cent of the respondents filled in the complete questionnaire without missing data and after giving their informed consent. The sample size was considered as acceptable based on the common rule of thumb of 10:1 person-to-item ratio.

Table 1: Q4TE items and Italian translation

Short- or long-term evaluation (Wang & Wilcox, 2006)	Four-level evaluation (Kirkpatrick, 1967)	Original Q4TE scale (Grohmann & Kauffeld, 2013)	Item wording (Italian and English version)
<i>Short-term evaluation</i>	<i>Reaction</i>	<i>Satisfaction</i>	<i>Ho apprezzato molto il corso di formazione</i> I enjoyed the training very much <i>Conservero un buon ricordo del corso di formazione</i> I will keep the training in good memory
		<i>Utility</i>	<i>Il corso di formazione ha apportato un grande beneficio al mio lavoro</i> The training is very beneficial to my work <i>La partecipazione a questo tipo di formazione è molto utile per il mio lavoro</i> Participation in this kind of training is very useful for my job
<i>Long-term evaluation</i>	<i>Learning</i>	<i>Knowledge</i>	<i>Dopo la formazione ho molte più conoscenze di prima rispetto ai contenuti del corso</i> After the training, I know substantially more about the training contents than before <i>Ho appreso molte cose nuove durante il corso di formazione</i> I learned a lot of new things in the training
		<i>Application to practice</i>	<i>Nel mio lavoro quotidiano, utilizzo spesso la conoscenza che ho acquisito durante il corso di formazione</i> In my everyday work, I often use the knowledge I gained in the training <i>Riesco bene ad applicare i contenuti del corso nel mio contesto lavorativo quotidiano</i> I successfully manage to apply the training contents in my everyday work.

Table 1: Continued

Short- or long-term evaluation (Wang & Wilcox, 2006)	Four-level evaluation (Kirkpatrick, 1967)	Original Q4TE scale (Grohmann & Kauffeld, 2013)	Item wording (Italian and English version)
	<i>Results</i>	<i>Individual organizational results</i>	<p><i>Dopo il corso di formazione, sono più soddisfatto del mio lavoro</i></p> <p>Since the training, I have been more content with my work</p> <p><i>La mia performance lavorativa è migliorata grazie all'applicazione dei contenuti del corso</i></p> <p>My job performance has improved through the application of the training contents</p>
		<i>Global organizational results</i>	<p><i>In generale, mi sembra che l'applicazione dei contenuti del corso abbia facilitato il lavoro nella mia organizzazione</i></p> <p>Overall, it seems to me that the application of the training contents has facilitated the work flow in my company</p> <p><i>In generale, mi sembra che il clima organizzativo sia migliorato grazie al corso di formazione</i></p> <p>Overall, it seems to me that the organizational climate has improved due to the training</p>

### Measures

Q4TE (Grohmann & Kauffeld, 2013) is a 12-item questionnaire, measured on a 11-point Likert-type scale, with alternatives ranging from 0 ('completely disagree') to 10 ('completely agree'). The questionnaire consists of 12 items referring to the six sub-scales of satisfaction, utility, knowledge, application to practice, individual organizational results and global organizational results. Each sub-scale is composed of two items, which keeps the measure time-efficient and psychometrically valid (Rammstedt, 2007; Rammstedt & John, 2007).

### Data analyses

We inspected the distributional properties of the scale to check the normality of the total scores. Skewness and kurtosis values between  $-2$  and  $+2$  were assumed to

Table 2: Item factor loading after EFA (factor loadings refer to the Italian version)

Item	Factor	
	1	2
I enjoyed the training very much	-0.115	<b>-0.965</b>
I will keep the training in good memory	-0.079	<b>-0.964</b>
The training is very beneficial to my work	0.182	<b>-0.771</b>
Participation in this kind of training is very useful for my job	0.272	<b>-0.660</b>
After the training, I know substantially more about the training contents than before	0.328	<b>-0.571</b>
I learned a lot of new things in the training	0.338	<b>-0.564</b>
In my everyday work, I often use the knowledge I gained in the training	<b>0.665</b>	-0.278
I successfully manage to apply the training contents in my everyday work	<b>0.597</b>	-0.296
Since the training, I have been more content with my work	<b>0.656</b>	-0.253
My job performance has improved through the application of the training contents	<b>0.836</b>	-0.075
Overall, it seems to me that the application of the training contents has facilitated the work flow in my company	<b>0.971</b>	0.097
Overall, it seems to me that the organizational climate has improved due to the training	<b>0.896</b>	0.095

Note: Extraction method: maximum likelihood. Rotation method: Oblimin with Kaiser normalization. Factor loadings greater than 0.3 have been highlighted in bold for ease of identification.

indicate an acceptable range to prove normal univariate distribution (George & Mallery, 2010).

EFA with maximum likelihood (ML) was performed to extract underlying common variance among items. Each item was included in a specific factor if there was a minimal factor loading of 0.4; whereas, to determine how many factors should be retained, we used the Kaiser criterion including factors with an eigenvalue  $>1$ , screeplot analysis and Horn's parallel analysis.

To assess the sampling adequacy, Kaiser-Meyer-Olkin (KMO) was calculated with values ranging from 0.7 to 0.8 (Hutcheson & Sofroniou, 1999), indicating that the sampling is adequate. Anti-image correlations of  $>0.5$  were deemed as acceptable (Field, 2009).

## Results from Study 1

The KMO of 0.93 verified the sampling adequacy for the EFA. Anti-image correlation values for individual items were all  $>0.80$ , which is well above the acceptable limit of 0.50.

As shown in Table 2, overall factor loadings were satisfactory (from  $-0.56$  to  $0.97$ ), whereas communalities ranged from 0.69 to 0.83.

An EFA was used to test the dimensionality of the Q4TE. In the initial EFA (ML), two factors with eigenvalue greater than 1.0 were extracted which accounted for 79.42 per cent of the variance of the original items (Table 2).

Horn's parallel analysis suggested that only one factor should be retained (adjusted eigenvalue = 7.88), whereas screeplot analysis indicated three underlying factors. These contradictory results led us to test different competing models through CFA in Study 2.

## Methodology of Study 2

### Participants

A convenience sample of 122 participants was collected through an online survey respecting the same criteria used in Study 1. The sample was composed of 63.9 per cent women and 36.1 per cent men, with mean age of 37.7 years (SD = 11.25) and average tenure of 6.7 years (SD = 8.00).

One hundred per cent of the respondents completed the questionnaire and gave their informed consent. The sample size was considered acceptable as ensuring the person-to-item ratio of 10:1.

### Measures

Consistently with the original validation study (Grohmann & Kauffeld, 2013), transfer to practice was measured with the item 'Have you been able to transfer training contents to practice?', which had to be rated with a positive (yes) or negative (no) answer (adapted from Kauffeld *et al.*, 2008, 2009).

Similarly, we measured transfer quantity through an item concerning the number of training contents/skills transferred to practice (Kauffeld & Lehmann-Willenbrock, 2010; Kauffeld *et al.*, 2008). Participants were actually asked to write down up to 10 training contents they had been able to apply in their everyday practice (Grohmann & Kauffeld, 2013).

### Data analysis

In order to test the structure of the scale extracted from EFA but also to examine the different competing models, including those already considered in the original validation study, we ran a CFA with ML estimation. Given the small size of our sample, different components of fit were evaluated (Hu & Bentler, 1995) considering multiple measures to highlight different aspects of fit (Tanaka, 1993). The  $\chi^2$  ratio ( $\chi^2$ /degrees of freedom [df]) was used to evaluate stand-alone models. This index tends to be less sensitive to sample size, and values less than 3 are taken to indicate acceptable models (Kline, 2010). We also used the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR), as well as Confirmatory Fit Index (CFI; Bentler, 1990), Tucker–Lewis Index (TLI) and Akaike Information Criterion (AIC; Akaike, 1987) (Schumacker and Lomax, 2014). As Hair *et al.* (2014) indicated, no definitive rule can determine a good model of fit; therefore, a variety of indices observation is suggested especially in small samples.

Generally, smaller  $\chi^2$ /degrees of freedom values correspond to better fitting models and are less sensitive to sample size. Values of RMSEA and SRMR up to 0.08 are generally considered as indicating adequate fit, whereas CFI and TLI values greater than 0.95 are usually acceptable (Hu & Bentler, 1999).

The reliability of Italian version of the Q4TE was examined by using Spearman–Brown coefficient (split half) for the two-item subscales, as suggested by Eisinga *et al.* (2013).

To test the discriminant validity of the scales, a mean comparison was performed between respondents who were able to transfer the learning into practice against respondents who were not. In addition, we investigated the relationship between the Q4TE scales and transfer quantity through bivariate correlation.

Bivariate correlation was also used to investigate the relationship of Q4TE scales with gender, age, organizational tenure, type of training (closed, open skills or both), number of training methodologies used during the training (ranging from 1 to 5 methodologies to be chosen from: frontal lectures, experiential exercises, group and peer discussion, simulations, virtual learning sessions), training duration and time lag between training and survey.



## Results of Study 2

In the light of the contradictory results emerging from EFA in Study 1, we wanted to test different models, which had already been tested in the original validation study and which could result in a more consistent factorial structure.

We first tested a two inter-correlated factor model solution, in line with our EFA results and with Wang and Wilcox model (2006). We then tested a one-factor model, encompassing all the items in one sole dimension of learning transfer, as suggested by the results of Horn's parallel analysis. The three-factor solution suggested by our scree plot analysis was not performed because it was not in line with any theoretical model.

Goodness of fit indices for the first two models, though, did not result in satisfactory indices. Therefore, we examined the models that had already been tested in the original scale validation: a two-latent second-order inter-correlated factors model, a four-latent inter-correlated factors model following Kirkpatrick's four levels model and a six-latent first-order factors model (Grohmann & Kauffeld, 2013).

Among all the tested models, only the last one (Model 5) showed the best indices of fit, consistently with the original validation study. Actually, it had good  $\chi^2$  ratio ( $\leq 3$ ) and SRMR values (0.03). CFI and TLI were satisfactory as well ( $\geq 0.95$ ). RMSEA was the only indicator that showed slightly higher values than expected (0.12), albeit well below the RMSEA values of the other models (Table 3).

This final model was thus chosen for the Italian validation of the Q4TE, which is composed of six-latent inter-correlated factors referring to satisfaction, utility, knowledge, application to practice, individual organizational results and global organizational results.

High inter-correlations were found between all Q4TE scales, ranging from 0.60 (between satisfaction and global organizational results) and 0.93 (between individual organizational results and global organizational results).

Reliability analyses, performed through the Spearman-Brown coefficient using the split half method, showed high internal consistency ranging from 0.93 (UT, APP and G-OR scales) to 0.97 (SAT and I-OR scale) (Table 4).

Evidence of discriminant validity was assured by a significantly higher scores in all the six scales for respondents who said they had been able to transfer the learning into practice against respondents who were not, with  $p$  ranging from  $< 0.05$  in the knowledge scale (mean = 14.30, SD = 4.81 vs. mean = 11.87, SD = 4.88,  $t(121) = 2.17$ ,  $p < 0.05$ ) to  $p < 0.001$  in the individual organizational results and global organizational results scales (I-OR mean = 11.90, SD = 5.37 vs. mean = 7.48, SD = 4.50,  $t(121) = 3.60$ ,  $p = 0.001$ ; G-OR mean = 11.23, SD = 5.12 vs. mean = 7.36, SD = 4.63,  $t(121) = 3.26$ ,  $p = .001$ ).

The relationship with transfer quantity was investigated with Pearson's correlation coefficient. Even if to a modest extent, transfer quantity was positively correlated, respectively, with the knowledge, application to practice, individual organizational results and global organizational results scales (Table 5).

Potential differences between groups of respondents were investigated through bivariate correlations as well. Gender, age, organizational tenure, training duration, time lag between training and survey, number of training methodologies used during the training and type of training (open, closed or both) were taken into account. No significant relationship was found between the Q4TE scores and gender, training duration, time lag between training and survey and number of training methodologies used during the training.

However, some statistically significant, albeit weak, relationships were found between Q4TE scales and age and tenure. In particular, satisfaction, knowledge and application scales showed negative correlations with both age and organizational tenure.

Type of training also resulted in a positive statistically significant relationship, with open skills correlated with satisfaction and utility scales (Table 6).

## Discussion

In the present paper, we illustrated the psychometric properties of the Italian version of the Questionnaire for Professional Training (Q4TE), originally validated by Grohmann and Kauffeld (2013).

Table 3: CFA results with goodness of fit indices comparison in Study 2

Model	$\chi^2$	d.f.	$\chi^2 / \text{d.f.}$	RMSEA	SRMR	CFI	TLI
(1) Two, inter-correlated factor model (resulting from EFA in Study 1): Short-term evaluation with six items Long-term evaluation with six items	350.60	53	6.61	0.21	0.06	0.86	0.83
(2) One-factor model (resulting from Monte Carlo analysis in Study 1): 12 inter-correlated items	714.33	54	13.23	0.39	0.12	0.70	0.63
(3) Two-latent second-order inter-correlated factors and six first-order factors (as tested in original validation): Short-term evaluation: SAT with two items, UT with two items, KNOW with two items Long-term evaluation: APP with two items, G-IO with two items, G-OR with two items	672.99	52	13.00	0.31	0.09	0.72	0.64
(4) Four latent inter-correlated factors following Kirkpatrick's (1967) four-level model (as tested in original validation): Reaction with four items Utility with two items Behavior with two items Results with four items	256.91	48	5.35	0.19	0.08	0.90	0.81
(5) Six-latent first-order factors (as resulting from original validation): SAT with two items UT with two items KNOW with two items APP with two items I-OR with two items G-OR with two items	107.46	39	2.75	0.12	0.03	0.97	0.95

Note: SAT = satisfaction, UT = utility, KNOW = knowledge, APP = application to practice, I-OR = individual organizational results, G-OR = global organizational results.

Table 4: Q4TE scales inter-correlations (Pearson's R) and reliability (Spearman–Brown coefficient – split half)

Scales	SAT	UT	KNOW	APP	I-OR	G-OR
SAT	(0.97)					
UT	0.798**	(0.93)				
KNOW	0.806**	0.832**	(0.94)			
APP	0.648**	0.839**	0.713**	(0.93)		
I-OR	0.615**	0.752**	0.628**	0.877**	(0.97)	
G-OR	0.602**	0.735**	0.618**	0.856**	0.933**	(0.93)

Note: Internal consistency values calculated with Spearman–Brown coefficient are shown diagonally (in parentheses).

\*\* $p < 0.01$  (two-tailed).

SAT = satisfaction, UT = utility, KNOW = knowledge, APP = application to practice, I-OR = individual organizational results, G-OR = global organizational results.

As a short measurement tool, in line with the original authors, we can say that Q4TE provides several advantages: it encompasses both short-term and long-term training outcomes (Wang & Wilcox, 2006) as well as more specific evaluation levels (Kirkpatrick & Kirkpatrick, 2007), it is extremely time efficient, it is applicable to a wide variety of training contexts and aims – either referring to closed or open skill development – and it shows sound psychometric properties.

In translating the Italian version of the tool, we tried to maintain the same advantages listed above, with the aim to respond to the demand for valid and reliable evaluation instruments, which is constantly growing also among Italian L&D experts and practitioners.

We led two studies to explore the consistency and psychometric solidity of the Italian form of the questionnaire as well as to explore its potential links to other individual and training-related variables.

In line with the findings by Grohmann and Kauffeld (2013), in Study 1 we ran an EFA which led to a two-factor solution accounting for short- and long-term learning outcomes, which was consistent with Wang and Wilcox conceptual model (2006). Nevertheless, a further factor exploration highlighted the potential presence of only one general factor. This led us to examine different factorial structure models in Study 2.

Indeed, in Study 2 we performed a five-model comparison, which included: a two-inter-correlated factor model (resulting from EFA in Study 1), one-factor model (resulting from Horn's parallel analysis in Study 1), a two-latent second-order inter-correlated factors and six-first-order factors model (as tested in original validation), a four-latent inter-correlated factors model following Kirkpatrick's (1967) four-level model (as tested in original validation) and eventually a six-latent first-order factors model (as resulting from original validation).

Table 5: Correlations between transfer quantity and satisfaction, utility, knowledge, application to practice, individual organizational results and global organizational results (Pearson's R)

	SAT	UTI	KNOW	APP	I-OR	G-OR
Transfer quantity	0.161	0.177	0.209*	0.223*	0.263**	0.227*

Note: \* $p < 0.05$ ; \*\* $p < 0.01$  (two-tailed).

SAT = satisfaction, UT = utility, KNOW = knowledge, APP = application to practice, I-OR = individual organizational results, G-OR = global organizational results.

Table 6: Q4TE scales correlations with gender, age, organizational tenure, training methodologies and type of training (Pearson's R)

Scales	SAT	UT	KNOW	APP	I-OR	G-OR
Gender <sup>a</sup>	0.009	-0.038	-0.074	0.045	0.111	0.054
Age	-0.201*	-0.147	-0.269**	-0.198*	-0.169	-0.114
Organizational tenure	-0.262**	-0.139	-0.255**	-0.193*	-0.157	-0.105
No. of training methodologies	0.132	0.096	0.070	0.035	0.040	-0.003
Type of training <sup>b</sup>	0.303**	0.166*	0.069	0.095	0.134	0.127

Notes: <sup>a</sup>Gender: 1 = female, 2 = male.

<sup>b</sup>1 = closed skills, 2 = both open and closed skills and 3 = open skills. Kendall's *t* correlations were used as type of training content is an ordinal variable.

\* $p < 0.05$ ; \*\* $p < 0.01$  (two-tailed).

As occurred in the international validation of the Q4TE, although at a first stage two-factor solutions seemed to provide a coherent framework, our CFA found the best model fit in the six inter-correlated first-order factors model, respectively, composed of satisfaction, utility, knowledge, application to practice, individual organizational results and global organizational results scales, represented by two items each (Research Question 1). Such a differentiation of dimensions allows the measurement of specific training outcomes at various levels. However, in agreement with Grohmann and Kauffeld (2013), we conclude that, if needed, a conceptual aggregation of the six dimensions within short- and long-term outcomes could be useful and coherent with the identified model.

By further exploring Study 2 results, we analysed the capability of the questionnaire to be associated with other measures of learning transfer. In particular, we found that respondents who reported to be able to transfer the learning into their everyday work had significantly higher scores in all the Q4TE scales than respondents who were not able to do it (Research Question 2). Moreover, we found a significant and positive relationship between the quantity of training contents applied back at work and some of the Q4TE scales, namely the knowledge, application to practice, individual organizational results and global organizational results scales (Research Question 3). These results are also consistent with what already found by Grohmann and Kauffeld (2013) who discovered that Q4TE successfully differentiated between respondents who could transfer learning from those who could not and that only application to practice had a positive relationship with transfer quantity. In both cases, it appears that Q4TE scales can successfully discriminate between simple reactions or perception of usefulness and concrete implementation of new learning and behaviors acquired during the training. This behavioral outcome, more than reactions, is particularly important as it is what, ultimately, makes training effective (Aguinis & Kraiger, 2009; Bell *et al.*, 2017).

Reliability of the Q4TE, albeit calculated on two-item scales, proved to be high and in line with the original validation. In this regard, we chose to use the split-half method (Eisinga *et al.*, 2013), as it provided further solidity to the measurement of the scales internal consistency.

Eventually, we also investigated if the Q4TE scores could be associated with other individual or training-related variables (Research Question 4). Age and organizational tenure showed a significant inverse relationship with Q4TE scores, highlighting how learning transfer can be stronger in a younger and more junior population. This result, albeit just hinted, seems to suggest that training can represent a useful development opportunity, which can be exploited at the most when expertise and seniority are less present but motivation to transfer is higher (Massenberg *et al.*, 2017).

Training type can also play a role in learning transfer, as interventions aimed at open skills development seem to lead to higher learning transfer. In line with previous pieces

of research, this may be due to the possibility – in open skills – to have a wider variety of contexts to transfer learning (Baldwin *et al.*, 2009; Blume *et al.*, 2010).

## Limitations and further research

Among the several limitations of this study, we should list limits which are intrinsic to the Q4TE itself and limits which are more specifically related to our study. With regard to the first type of limit, as also Grohmann and Kauffeld (2013) highlighted, the current study relied completely on computer-based, cross-sectional, retrospective samples. The simultaneous administration of measures referring to different training evaluation levels (Kirkpatrick & Kirkpatrick, 2007) can actually lead to different outcomes, as highlighted in other previous meta-analyses (see Alliger *et al.*, 1997). This limitation needs to be addressed with further research aiming at measuring the different outcomes and levels at different times.

Second, the Q4TE is a self-report scale, which can lead to several types of response bias (Podsakoff *et al.*, 2012). This is particularly true when assessing outcomes referring to behavior (application to practice) and organizational results (individual or global). This limit could be overcome with the use of more objective measurement methods, such as performance indicators, success cases collection and ROI estimates (Brinkerhoff, 2006; Kirkpatrick & Kirkpatrick, 2016; Phillips, 2007).

Nonetheless, this scale is capable of providing a time- and cost-efficient measure of outcomes which would be more expensive to identify through other methods.

Finally, the Q4TE six scales are made of two items only. Although some researchers do not support this type of scale construction (Credé *et al.*, 2012; Eisinga, Grotenhuis, & Pelzer, 2013; Hinkin, 1998; Loo, 2002), such a short measure can evidently provide a time-saving, practical tool that can be easily used to monitor the learning outcomes of large-scale or low-budget training initiatives, which cannot benefit from big investments on evaluation.

About our specific study, we should acknowledge some further limitations. The first one is related to the sample size, which is small and cannot represent at best the variety of training recipients to whom the Q4TE could be addressed. Further research with a wider sample could be very advantageous, both for the representativeness of the study and for the psychometric properties of the scale itself.

With reference to this, a second limit needs to be addressed. Although the six-factor model in our CFA was in line with the results of the Q4TE original validation and did show the best fit indices compared to the other competing models, we are aware that RMSEA may not be satisfactory from a purely statistical point of view, because values  $>0.1$  are generally seen as indicating poor fit. Although goodness of fit indices cannot be represented by fixed values defined once for all (Hair *et al.*, 2014) and, in particular, an absolute value for RMSEA is quite debated in the literature (Cangur & Ercan, 2015; Chen *et al.*, 2015; Rigdon, 1996), we strongly invite other researchers to further test the Italian version of the Q4TE by enlarging the sample and checking the factorial structure found in this study, in order to strengthen the soundness of the Italian version of the tool.

Another limit that we would like to underline is the lack of additional concurrent measures, more related to contextual factors (and not just individual ones), which could shed light on the complex interaction existing between a training intervention, its recipients, its designers, the work context hosting it and the relational, cultural or even symbolic dimensions (Caputo & Langher, 2015; Langher *et al.*, 2014) which enable its effectiveness within an organization or institution.

Further research could include the observation of other variables related to individual, training-related and organizational dimensions (Baldwin & Ford, 1988; Bocciardi *et al.*, 2017; Diamantidis & Chatzoglou, 2014; Holton, 2005; Kontoghiorghes, 2001) which may reduce, enhance or interact with training outcomes and potentially also with Q4TE scores. With reference to this, also in response to the different limitations already identified, further studies could benefit from longitudinal designs, which can conceptualize the training and its evaluation as a multifactorial process influenced also

by time and by what precedes and follows the classroom sessions (Argentin *et al.*, 2014).

## Implications for practice

In agreement with the Q4TE authors, we propose this tool as a very agile measurement instrument capable of collecting comparable results from several training interventions and settings, relying on a very well-known model which is widely used in training and L&D practice (Kirkpatrick & Kirkpatrick, 2007; Wang & Wilcox, 2006).

Being the Q4TE usable for different types of training in different kinds of sectors and industries, it is evident its applicability and capability to provide a benchmark for diverse sorts or stages of training interventions.

In addition, differently from English speaking countries, in Italy we do not have access to many other valid learning transfer measurement tools. Therefore, the Italian version of the scale represents an important first step in this direction.

Eventually, the short nature of the tool can definitely allow the combination with other quantitative (Phillips, 2012) and qualitative measures (Guerci & Vinante, 2011), aiming at both practice or research goals.

In a challenging market in which L&D investments become more and more limited and accurate, having a low cost but yet valid and fast evaluation measure for professional training can represent a great opportunity to spread a culture of intervention assessment and improvement. In such a context, the Italian version of the Q4TE is valuable and usable tool, capable to fill the current vacuum of evaluation instruments and practices among several L&D practitioners and experts.

## References

- Aguinis, H. and Kraiger, K. (2009), 'Benefits of training and development for individuals and teams, organizations, and society', *Annual Review of Psychology*, **60**, 1, 451–74.
- Akaike, H. (1987), 'Factor analysis and the AIC', *Psychometrika*, **52**, 3, 317–32.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H. and Shotland, A. (1997), 'A meta-analysis of the relations among training criteria', *Personnel Psychology*, **50**, 2, 341–58.
- Alvarez, K., Salas, E. and Garofano, C. M. (2004), 'An integrated model of training evaluation and effectiveness', *Human Resource Development Review*, **3**, 4, 385–416.
- Andrews, D., Nonnecke, B. and Preece, J. (2003), 'Electronic survey methodology: a case study in reaching hard-to-involve Internet users', *International Journal of Human-Computer Interaction*, **16**, 2, 185–210.
- Argentin, G., Pennisi, A., Vidoni, D., Abbiati, G. and Caputo, A. (2014), 'Trying to raise (low) math achievement and to promote (rigorous) policy evaluation in Italy: evidence from a large-scale randomized trial', *Evaluation Review*, **38**, 2, 99–132.
- ASTD. (2009), *The value of evaluation: making training evaluations more effective*. American Society of Training and Development Report Executive Summary. Available at <https://www.td.org/research-reports/astd-research-the-value-of-evaluation>
- Baldwin, T. T. and Ford, J. K. (1988), 'Transfer of training: a review and directions for future research', *Personnel Psychology*, **41**, 1, 63–105.
- Baldwin, T. T., Ford, J. K. and Blume, B. D. (2009), 'Transfer of training 1988–2008: an updated review and agenda for future research', *International Review of Industrial and Organizational Psychology*, **24**, 1, 41–70.
- Bates, R., Holton, E. F. and Hatala, J. P. (2012), 'A revised learning transfer system inventory: factorial replication and validation', *Human Resource Development International*, **15**, 5, 549–69.
- Bell, B. S., Tannenbaum, S. I., Ford, J. K., Noe, R. A. and Kraiger, K. (2017), '100 years of training and development research: what we know and where we should go', *Journal of Applied Psychology*, **102**, 3, 305–23.
- Bentler, P. M. (1990), 'Comparative fit indexes in structural models', *Psychological bulletin*, **107**, 2, 238.
- Blume, B. D., Ford, J. K., Baldwin, T. T. and Huang, J. L. (2010), 'Transfer of training: a meta-analytic review', *Journal of Management*, **36**, 4, 1065–105.
- Boccardi, F., Caputo, A., Fregonese, C., Langher, V. and Sartori, R. (2017), 'Career adaptability as a strategic competence for career development: an exploratory study of its key predictors', *European Journal of Training and Development*, **41**, 1, 67–82.

- Brinkerhoff, R. O. (2006), *Telling Training's Story: Evaluation Made Simple, Credible, and Effective*. Business Ebooks Online (Oakland, CA: Berrett-Koehler Publishers).
- Cangur, S. and Ercan, I. (2015), 'Comparison of model fit indices used in structural equation modeling under multivariate normality', *Journal of Modern Applied Statistical Methods*, **14**, 1, 14.
- Caputo, A. and Langher, V. (2015), 'Validation of the collaboration and support for inclusive teaching scale in special education teachers', *Journal of Psychoeducational Assessment*, **33**, 3, 210–22.
- Chen, F., Curran, P. J., Bollen, A. K., Kirby, J. and Paxton, P. (2015), 'An empirical evaluation of the use of fixed cutoff points in the RMSEA test statistic in structural equation models', *Sociological Methods and Research*, **36**, 4, 462–94.
- Credé, M., Harms, P., Niehorster, S. and Gaye-Valentine, A. (2012), 'An evaluation of the consequences of using short measures of the Big Five personality traits', *Journal of Personality and Social Psychology*, **102**, 4, 874–88.
- Diamantidis, A. D. and Chatzoglou, P. D. (2014), 'Employee post-training behaviour and performance: evaluating the results of the training process', *International Journal of Training and Development*, **18**, 3, 149–70.
- Eisinga, R., Grotenhuis, M. and Pelzer, B. (2013), 'The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?', *International Journal of Public Health*, **58**, 4, 637–42.
- Evans, J. and Mathur, A. (2005), 'The value of on-line surveys', *Internet Research*, **15**, 2, 2195–219.
- Field, A. (2009), *Discovering Statistics Using SPSS* (London: Sage).
- George, D. and Mallery, P. (2010), *SPSS for Windows Step by Step: A Simple Study Guide and Reference, 17.0 Update* (Boston, MA: Allyn & Bacon).
- Grohmann, A. and Kauffeld, S. (2013), 'Evaluating training programs: development and correlates of the questionnaire for professional training evaluation', *International Journal of Training and Development*, **17**, 2, 135–55.
- Grohmann, A., Beller, J. and Kauffeld, S. (2014), 'Exploring the critical role of motivation to transfer in the training transfer process', *International Journal of Training and Development*, **18**, 2, 84–103.
- Guerci, M. and Vinante, M. (2011), 'Training evaluation: an analysis of the stakeholders' evaluation needs', *Journal of European Industrial Training*, **35**, 4, 385.
- Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2014), *Multivariate Data Analysis: Pearson New International Edition. Always Learning* (Harlow Essex: Pearson).
- Hinkin, T. R. (1998), 'A brief tutorial on the development of measures for use in survey questionnaires', *Organizational Research Methods*, **1**, 1, 104–21.
- Holton, E. F. (2005), 'Holton's evaluation model: new evidence and construct elaborations', *Advances in Developing Human Resources*, **7**, 1, 37–54.
- Holton, E. F., Bates, R. A. and Ruona, W. E. (2000), 'Development of a generalized learning transfer system inventory', *Human Resource Development Quarterly*, **11**, 4, 333.
- Holton, E. F., Bates, R. A., Seyler, D. L. and Carvalho, M. B. (1997), 'Toward construct validation of a transfer climate instrument', *Human Resource Development Quarterly*, **8**, 2, 95–113.
- Hu, L.-T. and Bentler, P. M. (1995), 'Evaluating model fit', in Hoyle, R.H. (ed.), *Structural equation modeling: Concepts, issues, and applications*. (Thousand Oaks (CA): Sage), pp. 76–99.
- Hu, L. T. and Bentler, P. M. (1999), 'Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives', *Structural Equation Modeling: A Multidisciplinary Journal*, **6**, 1, 1–55.
- Hutcheson, G. D. and Sofroniou, N. (1999), *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models* (Thousand Oaks, CA: Sage).
- Kauffeld, S. and Lehmann-Willenbrock, N. (2010), 'Sales training: effects of spaced practice on training transfer', *Journal of European Industrial Training*, **1**, 23–37.
- Kauffeld, S., Bates, R., Holton Iii, E. F. and Müller, A. C. (2008), 'The German version of the Learning Transfer Systems Inventory (GLTSI): psychometric validation', *Zeitschrift für Personalpsychologie*, **7**, 2, 50–69.
- Kauffeld, S., Brennecke, J. and Strack, M. (2009), 'Erfolgsichtbarmachen: das Maßnahmen-Erfolgs-Inventar (MEI) zur Bewertung von Trainings [Visualizing Training Outcomes: The MEI for Training Evaluations]', in S. Kauffeld, S. Grote and E. Frieling (eds), *Handbuch Kompetenzentwicklung* (Stuttgart, Germany: Schäffer-Poeschel), pp. 55–78.
- Kirkpatrick, D. L. (1967), 'Whose responsibility is training?', *Training and Development Journal*, **21**, 3, 22.
- Kirkpatrick, D. L. and Kirkpatrick, J. D. (2007), *Implementing the Four Levels: A Practical Guide for Effective Evaluation of Training Programs* (San Francisco, CA: Berrett-Koehler Publishers).
- Kirkpatrick, J. D. and Kirkpatrick, W. K. (2016), *Kirkpatrick's Four Levels of Training Evaluation* (Alexandria, VA: Association For Talent Development).

- Kline, R. B. (2010), 'Promise and pitfalls of structural equation modeling in gifted research', in Thompson, B. & Subotnik, R.F. (eds.), *Methodologies for conducting research on giftedness* (Washington, DC: American Psychological Association), pp. 147–169.
- Kontoghiorghes, C. (2001), 'Factors affecting training effectiveness in the context of the introduction of new technology—a US case study', *International Journal of Training and Development*, **5**, 4, 248–60.
- Kraiger, K. (2002), *Creating, Implementing, and Managing Effective Training and Development: State-of-the-Art Lessons for Practice* (San Francisco, CA: Jossey-Bass).
- Langher, V., Brancadoro, B., D'Angeli, M. and Caputo, A. (2014), 'Imagining future internship in professional psychology: A study on university students' representations', *Universitas Psychologica*, **13**, 4, 1589–601.
- Langher, V., Caputo, A. and Ricci, M. E. (2017), 'The potential role of perceived support for reduction of special education teachers' burnout', *International Journal of Educational Psychology*, **6**, 2, 120.
- Lawshe, C. H. (1975), 'A quantitative approach to content validity', *Personnel Psychology*, **28**, 563–75.
- Lewis, T. (1996), 'A model for thinking about the evaluation of training', *Performance Improvement Quarterly*, **9**, 1, 3–22.
- Linton, M. (1982), 'Transformations of Memory in Everyday Life', in U. Neisser (ed.), *Memory Observed: Remembering in Natural Contexts* (San Francisco, CA: Freeman), pp. 77–91.
- Loo, R. (2002), 'A caveat on using single-item versus multiple-item scales', *Journal of Managerial Psychology*, **17**, 1, 68–75.
- Lourenço, D. F. M. E. C. (2016), *Self-regulated Learning and Training Effectiveness* (Lisboa, Portugal: ISCTE - Instituto Universitário de Lisboa).
- Massenberg, A. C., Schulte, E. M. and Kauffeld, S. (2017), 'Never too early: learning transfer system factors affecting motivation to transfer before and after training programs', *Human Resource Development Quarterly*, **28**, 1, 55–85.
- Massenberg, A.-C., Spurk, D. and Kauffeld, S. (2015), 'Social support at the workplace, motivation to transfer and training transfer: a multilevel indirect effects model', *International Journal of Training and Development*, **3**, 161.
- McKenzie, J. F., Wood, M. L., Kotecki, J. E. and Clark, J. K. (1999), 'Establishing content validity: using qualitative and quantitative steps', *American Journal of Health Behavior*, **23**, 4, 311–8.
- Nickols, F. W. (2005), 'Why a stakeholder approach to evaluating training', *Advances in Developing Human Resources*, **7**, 1, 121–34.
- Phillips, J. J. (2012), *Handbook of Training Evaluation and Measurement Methods. Building Blocks of Human Potential* (Hoboken, NJ: Routledge).
- Phillips, J. J. (2012), *Return on investment in training and performance improvement program*. (New York: Routledge).
- Phillips, J. J., and Phillips, P. P. (2016), *Handbook of training evaluation and measurement methods*. (New York: Routledge).
- Phillips, P. P., Ray, R. and Phillips, J. J. (2015), *Measuring the Success of Leadership Development: A Step-by-Step Guide for Measuring Impact and Calculating ROI* (Alexandria, VA: Association For Talent Development).
- Phillips, P. W. (2007), *The ROI Fieldbook: Strategies for Implementing ROI in HR and Training/Patricia Pulliam Phillips [et al]* (Amsterdam, Netherlands: Elsevier Butterworth-Heinemann).
- Pineda, P. (2010), 'Evaluation of training in organisations: a proposal for an integrated model', *Journal of European Industrial Training*, **34**, 7, 673–93.
- Podsakoff, P. M., MacKenzie, S. B. and Podsakoff, N. P. (2012), 'Sources of method bias in social science research and recommendations on how to control it', *Annual Review of Psychology*, **63**, 539–69.
- Rammstedt, B. (2007), 'The 10-item Big Five Inventory (BFI-10): norm values and investigation of socio-demographic effects based on a German population representative sample', *European Journal of Psychological Assessment*, **23**, 3, 193–201.
- Rammstedt, B. and John, O. P. (2007), 'Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German', *Journal of Research in Personality*, **41**, 1, 203–12.
- Rekik, G. and Bali, N. (2017), 'Tunisian physical education student trainees' agreement rate about the consistency between initial training and integration during the preparatory internship for professional life', *Creative Education*, **8**, 03, 373.
- Rigdon, E. E. (1996), 'CFI versus RMSEA: a comparison of two fit indexes for structural equation modeling', *Structural Equation Modeling*, **3**, 4, 369–79.
- Ritzmann, S., Hagemann, V. and Kluge, A. (2014), 'The Training Evaluation Inventory (TEI)-evaluation of training design and measurement of training outcomes for predicting training success', *Vocations and Learning*, **7**, 1, 41–73.



- Schumacker, R. E. and Lomax, R. G. (2004), *A beginner's guide to structural equation modeling*. (Psychology Press).
- Sitzmann, T. and Weinhardt, J. M. (2017), 'Approaching evaluation from a multilevel perspective: a comprehensive analysis of the indicators of training effectiveness', *Human Resource Management Review*.
- Sousa, V. D. and Rojjanasrirat, W. (2011), 'Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline', *Journal of Evaluation in Clinical Practice*, **17**, 2, 268–74.
- Swanson, R. A. (2007), *Analysis for Improving Performance: Tools for Diagnosing Organizations and Documenting Workplace Expertise. A Publication in the Berrett-Koehler Organizational Performance Series* (San Francisco, CA: Berrett-Koehler Publishers).
- Tanaka, J. S. (1993), 'Multifaceted Conceptions of Fit in Structural Equation Models', in A. Kenneth and L. J. S. Bollen (eds), *Testing Structural Equation Models* (Newbury Park, CA: Sage), pp. 10–39.
- Wang, G. G. and Spitzer, D. R. (2005), 'Advances in HRD measurement and evaluation: theory and practice', *Academy of Human Resource Development*, **7**, 1, 5–15.
- Wang, G. G. and Wilcox, D. (2006), 'Training evaluation: knowing more than is practiced', *Advances in Developing Human Resources*, **8**, 4, 528–39.
- Yelon, S. L. and Ford, J. K. (1999), 'Pursuing a multidimensional view of transfer', *Performance Improvement Quarterly*, **12**, 3, 58–78.