Iran Agricultural Research (2026) 45(2) 103-114

Shiraz University

Iran Agricultural Research

Journal homepage: https://iar.shirazu.ac.ir/?lang=en



Research Article

Determining expert-derived weights for Kirkpatrick's evaluation model levels and effectiveness grading cut-off points in assessing skill-based agricultural extension training courses

Arman Bakhshi Jahromi*

Socio-Economic and Agricultural Extension Research Department, Fars Agricultural and Natural Resources Research and Education Center, AREEO, Shiraz, I. R. Iran

ARTICLE INFO

Keywords:
Agricultural extension
Grading cut-offs
Kirkpatrick
Training effectiveness
Weighting

Received: 04 June 2026
Revised: 24 July 2026
Accepted: 27 July 2026

ABSTRACT- Kirkpatrick's four-level model (reaction, learning, behavior, and results) is the most widely used framework for training effectiveness evaluation. In recognizing a lack of an evidence-based weighting framework for Kirkpatrick's four evaluation levels, this study developed a practical weighting model based on expert consensus to improve the evaluation of agricultural extension training programs. This study aimed to determine expert-derived weights for the four levels of Kirkpatrick's evaluation model and to establish cut-off points for the effectiveness grade of the composite effectiveness score for agricultural extension training courses. As a quantitative survey study, the target population consisted of Ph.D. specialists in agricultural extension and education working in universities, the Agricultural Research, Education and Extension Organization (AREEO), and Agricultural Extension Coordination Offices. A researcher-developed questionnaire was sent to 64 specialists. Fifty-one complete responses were received (response rate 79.7%). Data were analyzed using descriptive statistics (frequency, percentage) and inferential statistics (Pearson's chi-square test). The findings showed no significant difference among university professors and AREEO personnel. The two groups had homogeneous opinions. Based on the weighted average method, the adjusted recommended weights for reaction, learning, behavior, and results were 15%, 20%, 35%, and 30%, respectively. The proposed effectiveness grading cut-offs were: score > 80 (high), 60–80 (moderate), 40–60 (weak), and < 40 (non-effective), which fully matched the specialists' modal choice. The results of this study can provide a layout for standardizing effectiveness evaluation in agricultural extension training courses in Iran.

INTRODUCTION

Training effectiveness evaluation is a systematic process for determining the extent to which training programs achieve their intended objectives and contribute to improvements in participants' knowledge, skills, behaviors, and organizational outcomes. In agricultural extension systems, where training serves as a key mechanism for enhancing farmers' competencies, promoting technology adoption, and supporting rural development, rigorous evaluation is essential for ensuring the efficient use of educational resources and informing evidence-based decision-making. However, training outcomes extend beyond immediate participant satisfaction to encompass changes in learning, workplace behavior, and long-term impacts. Therefore, their assessment requires multidimensional evaluation frameworks capable of capturing these interconnected dimensions. Accordingly, comprehensive models such as Kirkpatrick's four-level evaluation model have become widely recognized as effective frameworks for evaluating training effectiveness in both educational and extension contexts (Kirkpatrick & Kirkpatrick, 2016; Blume et al., 2010; Faisal-E-Alam & Islam, 2025). In recent years, the evaluation of agricultural training and extension courses has emerged as an important research area in agricultural education, rural development, and human capital management. The complexity of skill-based learning systems in agriculture, the need to transfer practical knowledge to farmers, and the necessity of measuring the real returns of training at the farm and rural community levels have encouraged researchers to adopt multidimensional evaluation models. Among these models, Kirkpatrick's four-level model (Kirkpatrick & Kirkpatrick, 2006; Kirkpatrick & Kirkpatrick, 2016), one of the most comprehensive, recognized, and widely used frameworks for training evaluation, has gained a particularly important position in studies of skill-based and extension training. Its four levels provide an appropriate framework for systematically assessing training outcomes across different dimensions, making the model well suited to research on the effectiveness of

*Corresponding Author: Assistant Professor, Socio-Economic and Agricultural Extension Research Department, Fars Agricultural and Natural Resources Research and Education Center, AREEO, Shiraz, I. R. Iran.
E-mail address: armanbakhshi@yahoo.com
DOI:10.22099/iar.2026.56316.1761

Copyright© 2026 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publisher.

agricultural training and extension courses. These four levels are:

Reaction (R): Trainees' satisfaction with training components (instructor, content, time, place, etc.).

Learning (L): Increase in trainees' knowledge, attitudes, and skills.

Behavior (B): Application of learned knowledge and skills in the work environment (e.g., using micronutrients in the field).

Results (O): Achievement of main goals and usefulness of training in work (e.g., increased yield).

To calculate a composite effectiveness score, the raw score for each level (0–100) is multiplied by its corresponding weight, and the resulting products are summed. However, a major challenge in applying Kirkpatrick's model is the absence of a standard and universal weighting system for the four levels. Researchers assign different weights depending on the type of training, target population, and objectives of the study. Although the model is widely accepted, there is no standardized procedure for weighting the four levels and aggregating their scores into a final effectiveness index. Many studies assign equal weights to all four levels, whereas recent evidence indicates that the relative importance of each level varies according to the nature of the training program, development objectives, and learner characteristics (Bakhshi Jahromi & Heydari, 2022). In the absence of an absolute criterion for determining weights, two main approaches can be distinguished. The first is a subjective, theory-based approach, which relies on methods such as expert panels or the assumption of equal weights. The second is an objective, data-driven approach, in which weights are derived through statistical and mathematical techniques such as the Analytic Hierarchy Process (AHP), principal component analysis (PCA), and entropy weighting (Paul et al., 2024). Thus, weighting may be based on findings from the literature, expert judgment, multi-criteria decision-making methods such as AHP, or other statistical and mathematical techniques. PCA, in particular, is widely used and accepted in agricultural education and generally assumes linear relationships among variables. Paul et al. (2024), for example, investigated agricultural training courses at Farmers' Training Centers (FTCs) in five Indian states and used a data-driven approach to calculate weights of 0.19, 0.23, 0.33, and 0.25 for reaction, learning, behavior, and results, respectively. These findings assigned the greatest weight to Level 3 (behavior), indicating that behavior was the most influential level in determining the overall effectiveness of agricultural training. More generally, weighting patterns reported in the literature tend to assign the lowest weight to reaction and greater weights to behavior and results (Reddy & Kumar, 2020). This pattern is consistent with the perspective of Kirkpatrick and Kirkpatrick (2016), who emphasize the greater importance of higher-level outcomes. However, not all studies follow this pattern. For example, Song (2024) proposed weights of 0.20, 0.30, 0.25, and 0.25 for reaction, learning, behavior, and results, respectively. Although these weights differ from the relative emphasis suggested by the originators of the model, they were derived in the context of a large industrial manufacturing company. The nature of industrial manufacturing training differs substantially from that of agricultural training, particularly in terms of work environments, skill requirements, technology adoption, and the transfer of learning to farm-level practice. Therefore, differences in weighting across studies may partly reflect differences in the contexts and objectives of the training programs being evaluated.

In agricultural extension training, determining the relative weight of each level of Kirkpatrick's model is particularly important because the ultimate goal of these courses is not merely to transfer knowledge but also to promote sustainable behavioral change on farms and improve the productive and economic performance of farmers and producers. Therefore, establishing contextually appropriate, expert-derived weights for the four levels can play an important role in providing a realistic assessment of overall training effectiveness. Kirkpatrick's model, which has been a dominant framework for training evaluation in organizational and educational settings since the 1950s, assumes that genuine training effectiveness should extend beyond initial participant satisfaction to encompass behavioral change and practical results. This perspective is particularly relevant to skill-based agricultural training, in which the success of training ultimately depends on whether farmers apply the knowledge and skills acquired during training and whether such application leads to observable production or economic outcomes. Recent evidence from agricultural training supports this view, indicating that the third and fourth levels are more important than the first two (Paul et al., 2024). Kirkpatrick and Kirkpatrick (2016) similarly emphasize the importance of Level 3 (behavior), identifying it as the most important level and Level 1 (reaction) as the least important. Several approaches can be used to determine the relative weights of the four levels. Some researchers assign equal weights of 25% to each level, whereas others derive weights from expert judgments (Praslova, 2010). Two approaches are particularly relevant to the present context: equal weighting and expert-based weighting.

Equal weighting: The simplest method for aggregating the four Kirkpatrick levels is to calculate their arithmetic mean and assign an equal weight of 25% to each level (Praslova, 2010). Although this approach is straightforward and computationally simple, it has been criticized on methodological grounds because it assumes that all four levels are equally important. In skill-based agricultural training, this assumption is often unrealistic because behavioral change and productive or economic results are generally more consequential than participants' initial satisfaction with the training.

Expert-based weighting: Another widely used approach is to determine weights through expert judgment, using techniques such as the AHP, the Delphi method, or pairwise comparisons. Experts in agricultural education, extension, and training evaluation assess the relative importance of the four levels based on their knowledge of the training context, objectives, and expected outcomes. The principal advantage of this approach is its flexibility and its ability to incorporate contextual knowledge that may not be captured by purely statistical methods. Recent research on agricultural

training evaluation indicates that experts tend to assign greater importance to Levels 3 (behavior) and 4 (results), because these levels are more directly associated with changes in farm performance and farmer productivity (Paul et al., 2024).

Statistical weighting: In recent years, statistical methods such as PCA, structural equation modeling (SEM), and factor analysis have increasingly been used to identify data-driven weights or assess the relative contribution of different indicators to a composite measure. Sathishwaran et al. (2025) examined the structural relationships among Kirkpatrick's four levels using SEM and found that the direct and indirect effects of the different levels on final training outcomes vary. Although their study did not directly use SEM to derive weights, these findings provide a methodological basis for assigning unequal importance to the four levels. In studies whose objective is to construct a composite effectiveness index, SEM can potentially provide a basis for weighting through standardized path coefficients, factor loadings, or total effects, depending on the structure of the model and the theoretical rationale for the index. Similarly, Paul et al. (2024) used PCA to derive indicator weights, demonstrating how data-driven methods can provide an alternative to subjective weighting. However, calculating a composite training effectiveness index raises another important question: What does the resulting score actually mean? For example, if the effectiveness score of a medicinal plant cultivation course is 67 on a 0–100 scale, it is necessary to determine whether 67 represents low, moderate, or high effectiveness. Without an interpretive framework, the numerical score alone provides limited practical information. Therefore, after calculating the final effectiveness score, researchers generally need to classify scores into meaningful categories or effectiveness levels. Such classification not only facilitates the interpretation of results but also supports managerial decision-making, comparison among courses, and prioritization of training programs. Several scientific approaches can be used to classify training effectiveness levels. One important basis is educational scaling theory, which provides a foundation for converting raw scores into interpretable categories in educational and psychometric assessment (Allen & Yen, 2002; Cohen & Swerdlik, 2021). When educational evaluation scores are standardized on a 0–100 scale, the score range can be divided into several categories to make the level of performance or effectiveness easier to interpret. The boundaries of these categories can be established using different approaches, two of which are particularly relevant to training evaluation: the criterion-referenced approach and the equal-interval approach.

Criterion-referenced approach: In many training effectiveness studies, classification is based on a criterion-referenced rather than a norm-referenced framework. Under this approach, category boundaries are established according to educational objectives and desired levels of performance rather than the performance of other participants (Nitko & Brookhart, 2014). For example, a score below 40 could indicate failure to achieve the intended educational objectives, a score above 60 could indicate partial achievement, and a score above 80 could indicate a highly desirable level of effectiveness. This approach is particularly applicable to skill-based and agricultural training because the primary objective is to assess the extent to which participants achieve the intended competencies and demonstrate the desired behavioral changes, rather than to rank individuals against one another.

Equal-interval approach: In this approach, the 0–100 score range is divided into intervals of equal or approximately equal width. For example, the range can be divided into five categories such as 0–20, 21–40, 41–60, 61–80, and 81–100, ranging from very poor to very strong effectiveness. The main advantage of this method is its simplicity and ease of interpretation. It also facilitates comparisons among courses or training programs by providing a common classification scale. However, because the category boundaries are determined mathematically rather than by reference to specific educational objectives, the resulting categories should be interpreted as descriptive levels rather than as direct evidence that a particular educational standard has or has not been achieved.

Mean and standard deviation classification: Some researchers use a statistical approach in which effectiveness levels are classified based on the mean and standard deviation of the observed scores. Under this approach, the overall mean effectiveness score is calculated, and categories are defined according to the distance of scores from the mean. For example, the middle category may include scores within ± 0.5 standard deviation of the mean, or, in some applications, within ± 1 standard deviation, with the remaining categories defined by adjacent ranges. This approach is based on z-scores and the properties of the normal distribution. Its main advantage is that the resulting classification reflects the actual distribution of the data. However, its main disadvantages are its dependence on the sample and the difficulty of making comparisons across studies. If the population or sample changes, the mean and standard deviation also change, and the same score may consequently be assigned to a different effectiveness category. Therefore, this approach is used less frequently in applied agricultural extension research. For example, Paul et al. (2024) used the mean and half of the standard deviation of aggregated weighted indices to classify the effectiveness of training courses into three categories: high, medium, and low (Table 1), assuming a normal distribution of the index values. This represents a data-driven approach because the classification thresholds are derived from the observed distribution of the effectiveness index. However, in many agricultural extension studies, only one or a small number of extension courses are evaluated, and sufficiently large datasets may not be available to establish reliable distribution-based categories. Therefore, for interpreting and classifying the composite effectiveness index in the present study, a percentage-based evaluation scale with predefined cut-off points is proposed.

Table 1. Effectiveness grading scale for training courses (Paul *et al.*, 2024)

Effectiveness score range	Effectiveness grade
$E > X + 0.5\sigma$	High
$\bar{X} - 0.5\sigma \leq E \leq \bar{X} + 0.5\sigma$	Medium
$E < \bar{X} - 0.5\sigma$	Low

Expert-based cut-off setting: In some studies, category boundaries are determined by specialists and subject-matter experts. In the educational measurement literature, this approach is known as standard setting and encompasses a range of methods for establishing performance standards and cut-off points (Tiyawongsuwan et al., 2013). The validity of expert-based standard setting depends strongly on the selection of appropriate experts, their adequate training, and the use of systematic and structured procedures. Under this approach, a panel of experts applies a defined standard-setting method to determine meaningful cut-off points. For example, experts may determine that a score below 50 represents non-effective training, whereas a score above 75 represents desirable effectiveness. This approach is commonly used in skill-based training, professional assessment, and policy research. Expert judgments can subsequently be summarized and analyzed using different procedures, including modal frequency or weighted scales.

Summary and research rationale: The evaluation of agricultural extension training courses continues to face methodological challenges because of the skill-based and applied nature of agricultural learning. Kirkpatrick's model, one of the most comprehensive and widely used frameworks for training evaluation, encompasses four levels: reaction, learning, behavior, and results. However, two fundamental problems remain in the practical application of this model within Iran's agricultural education system.

Problem 1: Lack of a standard and locally appropriate weighting system: There is no established standard or locally validated system for assigning weights to the four Kirkpatrick levels and aggregating them into a composite training effectiveness index. Researchers and extension specialists in Iran may either assign equal weights to all four levels, despite their potentially different importance, or assign weights arbitrarily without a clearly established methodological basis. Consequently, effectiveness scores obtained from different evaluations may not be directly comparable, limiting their usefulness for evidence-based managerial decision-making.

Problem 2: Lack of a standard grading scale: There is also no established standard grading scale or agreed set of cut-off points for interpreting the final composite effectiveness score. Without clearly defined thresholds, it is difficult to determine whether a given score represents low, moderate, or high training effectiveness. This limitation also complicates the comparison of courses and the prioritization of training programs for improvement or further investment. From a philosophical perspective, the present study adopts a pragmatic approach, recognizing that methodological decisions should be guided primarily by the nature of the research problem rather than by strict adherence to a single philosophical paradigm. While Kirkpatrick's evaluation model provides a comprehensive conceptual framework for assessing training effectiveness, it does not specify the relative importance of its four evaluation levels or provide a standardized procedure for converting their scores into a single composite index. Consequently, determining appropriate weights requires informed methodological and professional judgment rather than relying exclusively on purely objective measurement. In this context, expert knowledge represents an important source of evidence for developing a practical and theoretically grounded weighting frame. Incorporating the judgments of appropriately selected experts can help establish weights that reflect the relative importance of the four levels within the specific context of agricultural extension training. Similarly, expert-based standard setting can provide theoretically and contextually meaningful cut-off points for interpreting the resulting composite effectiveness index. Together, these approaches can improve the consistency, transparency, comparability, and practical applicability of training effectiveness evaluation in agricultural extension systems.

Recent studies have increasingly emphasized that training effectiveness should be evaluated using integrated and multidimensional frameworks rather than isolated indicators. Recent applications of Kirkpatrick's evaluation model have also highlighted the need for quantitative approaches that combine expert judgment with statistical techniques to improve the interpretability, consistency, and practical applicability of evaluation results (Faisal-E-Alam & Islam, 2025; Sathishwaran et al., 2025). Nevertheless, no standardized framework has yet been proposed for assigning expert-derived weights to Kirkpatrick's four evaluation levels or for establishing effectiveness-rating cut-off points specifically for agricultural extension training programs. Therefore, the central problem addressed in this research is the methodological gap resulting from the absence of a standard weighting and cut-off system for evaluating the effectiveness of Iranian agricultural extension courses. Given the importance of this issue, the present study addresses this gap by surveying specialists in agricultural extension and education in Iran and using a weighted-average method to aggregate their judgments. Unlike an approach based solely on the mode, the weighted-average method takes into account the full distribution of expert responses and therefore provides a more comprehensive representation of the collective judgment. Accordingly, the objectives of this study are to: 1. Determine the expert-derived weight assigned to each of Kirkpatrick's four evaluation levels. 2. Determine an appropriate grading scale and cut-off points for interpreting the composite training effectiveness score. 3. Investigate whether specialists' workplace or organizational setting influences their judgments regarding the relative importance of the four Kirkpatrick levels and the proposed effectiveness scale.

MATERIALS AND METHODS

This research employed a quantitative survey design using a researcher-developed questionnaire. The study population consisted of Ph.D. specialists in agricultural extension and education working at universities, the AREEO, and Agricultural Extension Coordination Offices. These groups were selected because of their pivotal roles in Iran's agricultural education and extension system. Participants were not randomly selected; rather, they were purposively recruited based on their relevant expertise and professional experience in agricultural extension and education.

Specifically, these groups were deliberately targeted because they represent major producers of scientific knowledge and key implementers of agricultural education and extension training programs. The questionnaire was developed as an online instrument and distributed through social media to 64 Ph.D. specialists in agricultural extension and education in Iran. A total of 51 complete responses were received, yielding a response rate of 79.7%. Among the respondents, 28 (54.9%) were university professors, 19 (37.3%) were personnel of AREEO, including faculty members, researchers, and instructors, and 4 (7.8%) were personnel of Agricultural Extension Coordination Offices. In expert-based survey research targeting Ph.D. specialists in a specific field, sample sizes are typically substantially smaller than those used in general population surveys. The recruitment of 64 experts was considered adequate for the present expert-based judgment study because the objective was not statistical inference from a representative population sample, but rather the elicitation of informed expert judgments and the identification of areas of consensus. Furthermore, the response rate of 79.7% indicates a high level of participant engagement and provides a strong basis for the analysis of the collected expert judgments. Previous methodological studies have emphasized that, in expert-based approaches, the quality and expertise of participants are more important than sample size (Okoli & Pawlowski, 2004; Hsu & Sandford, 2007). Similarly, studies involving the construction of composite indices have relied on expert judgment to assign indicator weights when no established objective weighting framework is available (te Lintelo et al., 2014).

The questionnaire was brief and consisted of four questions: one question on workplace; one five-option question concerning the weighting of Kirkpatrick's four evaluation levels (Table 1); one four-option question concerning the cut-off points for the effectiveness grading scale (Table 2); and one open-ended question inviting additional explanations or comments. The questionnaire was specifically developed to elicit expert judgments regarding the relative importance of Kirkpatrick's four evaluation levels and the corresponding effectiveness rating scale. Therefore, the primary methodological concern was to establish the content validity of the instrument rather than to assess its internal consistency. To establish content validity, the initial version of the questionnaire was reviewed by five specialists in agricultural extension and education. They assessed the relevance, clarity, comprehensiveness, and appropriateness of each question in relation to the study objectives. Their comments and suggestions were incorporated into the final version of the questionnaire before data collection. Because each expert judgment was obtained through a single item rather than a multi-item scale, internal consistency reliability indices, such as Cronbach's alpha, were not applicable. Such indices are designed for instruments containing multiple items that measure the same underlying construct. In contrast, the present questionnaire consisted of independent single-item questions designed to elicit expert judgments on distinct methodological issues. Accordingly, the quality of the instrument was established primarily through content validation by expert review.

Data were analyzed using SPSS version 26 using both descriptive and inferential statistical methods. Absolute and relative frequencies were calculated for each response option concerning the proposed weights and effectiveness grading scale, both for the total sample and by workplace. To examine the association between workplace and the selected weight option, as well as between workplace and the selected grading-scale option, Pearson's chi-square test was performed using 2 × 5 and 2 × 4 contingency tables, respectively. The significance level was set at 0.05. The Agricultural Extension Coordination Offices group (n = 4) was excluded from the chi-square analyses because its small sample size could violate the assumptions of the chi-square test. Two methods were used to determine the final weight assigned to each Kirkpatrick level. The first method was based on the modal response, in which the weight set receiving the highest frequency of specialist responses was selected as the preferred weighting scheme. In the second method, based on the weighted average, the final weight Ws for each level $s \in \{R, L, B, O\}$ was calculated according to Eq. (1).

$$W_S = \frac{\sum_{i=1}^{5} (f_i \times p_{i,s})}{N} \qquad \text{Eq. (1)}$$

where:

- fi = number of respondents who chose weight option i (i = A, B, C, D, and E)
- pi, s = percentage assigned to level s in option i
- N = total respondents (51)

Similarly, two methods were used to determine the cut-off points for the effectiveness grading scale. The first method was based on the modal response, in which the grading-scale option receiving the highest frequency of specialist responses was selected. In the second method, the weighted-average approach was used. Under this approach, the threshold (T) for each boundary between two adjacent effectiveness categories—High–Moderate, Moderate–Weak, and Weak–Non-effective—was calculated as follows:

$$\mathrm{T} = \frac{\sum_{j=1}^{4} (f_j \times t_j)}{N} \qquad \text{Eq. (2)}$$

where:

- fj = number of respondents who chose scale option j (j = A, B, C, and D)
- tj = numerical boundary value from that option (e.g., for High–Moderate: 90 for A, 85 for B, 80 for C, and 75 for D)
- N = 51

RESULTS AND DISCUSSION

Before analyzing the main results, it was necessary to determine whether the two respondent groups, i.e., university professors and AREEO personnel, differed significantly in their opinions regarding the weights assigned to the four Kirkpatrick levels and the proposed effectiveness grading scale. For this purpose, Pearson's chi-square test was used to compare the responses of the two groups. The results are presented in Table 2.

Table 2. Testing homogeneity of opinions between university professors and Agricultural Research, Education and Extension Organization (AREEO) personnel (chi-square)

Variable	χ^2	df	*p*-value	Interpretation
Level weight option	0.86	4	0.93	No significant relationship
Grading scale cut-off option	4.10	3	0.25	No significant relationship

The results show that the *p*-value for the weight options was 0.93, indicating no significant association between workplace and the choice of weight set. In other words, the responses and opinions of the two groups were similar. Accordingly, the responses of both groups were pooled and presented jointly. The primary objective of this study was to establish expert consensus (consensus building) for the standardization of the evaluation system, rather than to examine differences in experts' viewpoints. Consensus reflects the homogeneity and stability of expert opinions at the national level, which is highly valuable from a policy-making perspective because it facilitates the development of a single, unified standard. Therefore, the identification of consensus represents a strength of this research. Moreover, the *p*-value for the grading-scale cut-off option was 0.25, which was also non-significant. Therefore, there was no significant association between workplace and the choice of grading scale, with both groups predominantly expressing similar opinions. In summary, workplace (university versus AREEO) was not significantly associated with either the choice of level weights or the effectiveness grading scale. Thus, the specialists' opinions from both sectors were sufficiently homogeneous, and the aggregated results ($N = 51$) could be used as a basis for decision-making. The non-significant chi-square results provide evidence of agreement among the participating specialists and support the development of unified recommendations.

A: analysis of weights for Kirkpatrick's four levels

One of the most important recent developments in training evaluation is the move toward designing composite effectiveness indices. In this approach, the score of each Kirkpatrick level is multiplied by its corresponding weight, and the sum forms the final effectiveness index. The general formula for the final training effectiveness index, calculated as a weighted average of the four Kirkpatrick levels, can be expressed as:

$$\mathbf{E} = (W_R \times R) + (W_L \times L) + (W_B \times B) + (W_O \times O) \quad \text{Eq. (3)}$$

where:

- E = final training effectiveness index
- R, L, B, and O = scores for reaction, learning, behavior, and outcomes/results (each 0–100)
- W_R, W_L, W_B, and W_O = corresponding weights

All scores are in the range of 0–100. The main advantage of this approach is that it reflects the true importance of each level in the final evaluation. In agricultural extension training, this is especially important because some courses may generate high satisfaction but fail to change farmers' behavior or improve their performance. Hence, the question arises: what should be the weight of each level in the composite index for agricultural extension training courses? Table 3 shows the absolute and relative frequencies of the extension specialists' responses (N = 51) for each weight option. The sum of the weights for the levels in each option equals 100. Option D (10% reaction, 25% learning, 35% behavior, and 30% results) had the highest frequency (41.2%). Option E ranked second, Option B third, and Options A and C jointly fourth. The weighted average for the Kirkpatrick level weights (second method) was also calculated using Eq. (1), and the results are presented in Table 4. These results show minor differences from Option D, which was the modal choice. Although using weighted coefficients can increase evaluation accuracy, it also presents challenges. First, there is the subjectivity of expert judgment in determining the weights; second, the weights depend on the research context and the type of training program. The "proposed adjusted weights" are a combination of data, theory, and practicality: a theory-driven empirical adjustment. This adjustment both references theoretical foundations and shows that these weights do not deviate significantly from the collected data.

Table 3. Frequency distribution of weight options

Option	Level weight (%)					Frequency	Percent
	Reaction (R)	Learning (L)	Behavior (B)	Results (O)	Sum		
A	25	25	25	25	100	6	11.8
B	20	25	30	25	100	7	13.7
C	15	20	30	35	100	6	11.8
D	**10**	**25**	**35**	**30**	**100**	**21**	**41.2**
E	10	20	30	40	100	11	21.6
Total						51	100

Table 4. Weighted average calculation for each level

Level	$f_i \times p_{i,s}$	N	Weighted average (%)	Final rounded weight (%)
Reaction (R)	700	51	13.73	14
Learning (L)	1190	51	23.33	23
Behavior (B)	1605	51	31.47	32*
Results (O)	1605	51	31.47	31
Total	–	–	100	*To reach exact sum of 100, 1% added to Behavior (most important level per Kirkpatrick).

From a methodological perspective, the weights derived from the weighted average (14, 23, 32, and 31) provide the most precise reflection of the collective expert opinion. However, from a practical and theoretical perspective, it is sometimes necessary to adjust the weights for simplicity or better alignment with theory. Therefore, it is appropriate to align the research findings with the theoretical foundations of the model's originators. Accordingly, adjusted weights of 15%, 20%, 35%, and 30% are proposed (Table 5), and the scientific reasoning behind this adjustment is explained below.

Table 5. Kirkpatrick level weights based on different criteria

Level	Option D weight (%)	Weighted average (%)	Recommended weights (%)
Reaction (R)	10	14	15
Learning (L)	25	23	20
Behavior (B)	35	32	35
Results (O)	30	31	30

Theoretical foundation: Explicit priority of levels according to Kirkpatrick

Kirkpatrick and Kirkpatrick (2016) explicitly state that "Level 3 (behavior) is the most important level because, without behavioral change in the work environment, no real performance improvement occurs. On the other hand, Level 1 (reaction) is the least important level and should not be overemphasized because trainee satisfaction does not necessarily lead to performance change." This logic aligns with the philosophy of skill-based agricultural training. Taylor and Bhasme (2018) also state that behavioral change on the farm and increased yield are the ultimate goals of agricultural extension training. Accordingly, any valid weighting system must satisfy two basic conditions:

- Weight of behavior > Weight of other levels (the highest weight)
- Weight of reaction < Weight of other levels (the lowest weight)

All three weighting criteria in Table 5 satisfy these conditions, but in the weighted average, the gap between behavior and results is very small.

Alignment with expert opinions (research data)

The research data showed that in Option D (the most frequent), behavior weight is 35%. The proposed adjusted weights (15, 20, 35, and 30) are a combination of the modal option (D) and the weighted average:

- Behavior weight (35%) is exactly from Option D.
- Results weight (30%) is exactly from Option D.
- Reaction weight (15%) is close to the weighted average (14%) and also matches Option C (which had 12% of responses).
- Learning weight (20%) is close to the weighted average (23%) and matches Option E (20%) with over 21% of responses.

Practical-theoretical adjustment

In survey methodology, weights obtained from weighted averages (e.g., 14, 23, 32, and 31) may be unnecessarily precise for use in a simple composite formula and for communication with stakeholders. In such cases, a theory-driven empirical adjustment can be applied, whereby the empirically derived weights are rounded to practically meaningful values while remaining sufficiently close to the collective expert judgment. The proposed adjustment follows this principle. The reaction weight is adjusted from 14% to 15%, as this represents only a minor deviation from the empirical estimate while providing a simple and easily interpretable value. The learning weight is adjusted from 23% to 20%, consistent with the 20% learning weight assigned in both Options C and E, which together were supported by 33% of respondents. The behavior weight is increased from 32% to 35% to give greater emphasis to behavior, consistent with its theoretical priority and with the weighting assigned in the modal Option D. Finally, the results weight is adjusted from 31% to 30% to provide a simple and easily interpretable value while maintaining the required total of 100%. Thus, the proposed weights (15%, 20%, 35%, and 30%) represent a deliberate balance between empirical evidence, theoretical considerations, and practical usability.

Distance from weighted average (absolute error)

Calculating the absolute deviation (sum of the absolute differences) for the proposed adjusted weights shows that the distance is relatively small (8 percentage points), indicating that the proposed weights are close to the mean expert opinion:

Absolute error = $|15 - 14| + |20 - 23| + |35 - 32| + |30 - 31| = 1 + 3 + 3 + 1 = 8$

There is no universal absolute threshold for what constitutes a "small" error in such contexts. Therefore, describing an error of 8 percentage points as "relatively small" is a relative judgment. Considering the possible

range of weight changes (0 to 100 for each level), the maximum possible absolute difference between two weight sets (e.g., {100, 0, 0, 0} versus {0, 0, 0, 100}) is 200 percentage points. The error in this study is 8 percentage points; thus, the ratio of the observed error to the maximum possible error is 8/200 = 4%. From this perspective, an 8-percentage-point deviation can be considered "relatively small."

Sensitivity analysis

To ensure that this adjustment does not substantially alter the final evaluation of courses, a simple sensitivity analysis was performed. Assume a hypothetical course scores: reaction 77, learning 70, behavior 60, results 35. Calculating the composite effectiveness score with both weight sets:
- Weighted average weights (14, 23, 32, and 31)
- E = (0.14 × 77) + (0.23 × 70) + (0.32 × 60) + (0.31 × 35) = 10.78 + 16.10 + 19.20 + 10.85 = 56.93
- Proposed adjusted weights (15, 20, 35, and 30)
- E = (0.15 × 77) + (0.20 × 70) + (0.35 × 60) + (0.30 × 35) = 11.55 + 14.00 + 21.00 + 10.50 = 57.05

The difference is negligible (0.12 points). Thus, the sensitivity analysis shows that using the adjusted weights instead of the weighted average does not cause a meaningful change in the ranking of course effectiveness while respecting the theoretical framework. The proposed adjusted weights are considered "valid" because they: (1) adhere to theory, (2) have a small deviation from the experts' weighted average, (3) are rounded and easy to understand, ensuring high practicality, and (4) are supported by the opinions of agricultural extension and education specialists. These weights maintain the ranking structure derived from the data while ensuring full alignment with the theoretical foundations of the model. This indicates that the proposed adjustment improves theoretical consistency without materially affecting the composite effectiveness score.

B: Analysis of the grading cut-offs for effectiveness of skill-based agricultural extension training courses

After calculating the composite evaluation score, the next question is: what does this score indicate? How should the course be judged? Is it effective or not? If it is effective, to what degree? This is where the second objective of the study arises: determining the cut-offs for the effectiveness grades. Table 6 shows the frequency distribution of the scale options for effectiveness grading (range 0–100). The results show that, from the perspective of extension specialists, Option C (43.1%) is the most appropriate set of cut-offs for grading the effectiveness of agricultural skill-based training courses. Option D ranked second (37.3%), followed by Options B and A. Option A was very strict and was approved by only 4% of specialists. As previously mentioned, a second method based on the weighted average was also used to determine the effectiveness grading cut-offs. The results of the weighted average for the cut-offs (second method) are shown in Table 7. These cut-offs are highly consistent with Option C. Therefore, Option C is recommended as the effectiveness grading scale. The cut-offs (80, 60, and 40) are easy to understand, memorable, and compatible with conventional scoring systems. Recent research in agricultural extension training evaluation has moved toward developing quantitative and composite frameworks based on Kirkpatrick's model. Recent studies emphasize that using equal weights for the four levels cannot provide an accurate picture of actual training effectiveness. In contrast, the weighted average approach, based on expert opinion, statistical analysis, and structural modeling, allows for more precise measurement of training effectiveness. Moreover, using both subjective and data-driven methods can lead to more valid weights (Sathishwaran et al., 2025).

The findings of the present study show that Iranian agricultural extension and education specialists assign different weights to Kirkpatrick's four levels; specifically, they attribute greater importance to "behavior" (applying knowledge on the farm) and "results" (improving productive and economic performance) than to "reaction" and "learning." This finding is fully consistent with the perspective of the model's theorists (Kirkpatrick & Kirkpatrick, 2016), who identify Level 3 (behavior) as the most important and Level 1 (reaction) as the least important. The results also align with recent international studies, such as Paul et al. (2024), conducted in Indian Farmers' Training Centers, which reported weights of 0.19, 0.23, 0.33, and 0.25 for reaction, learning, behavior, and results, respectively. In both studies, behavior received the highest weight and reaction the lowest. The minor differences may be attributed to the differences in agricultural structures between Iran and India, differing development priorities, or differences in the composition of the expert panels. Song (2024), conducted in an industrial company, reported a different weighting pattern that contrasts with the present findings. This further confirms that the nature of training courses (agricultural versus industrial) can significantly affect the weighting of Kirkpatrick's levels.

In skill-based agricultural training, "behavior" and "results" usually have the greatest importance because the ultimate goal is to create lasting changes in productive and farm activities and, ultimately, to improve farmers' productivity and performance. However, it is also essential to note that Levels 3 (behavior) and 4 (results) are influenced not only by training but also by climatic, environmental, policy, and institutional support structures. In other words, achieving or not achieving Levels 3 and 4 is not solely dependent on educational factors. The influence of non-educational factors is much greater at Level 4 than at Level 3. Therefore, the weight assigned to behavior in agricultural skill-based training should be higher than that assigned to results. Moreover, in agricultural training, the primary goal is behavioral change among farmers because economic results are not sustainable without behavioral change. Therefore, designing an effectiveness index based on expert-derived weights can be a valuable tool for policymakers, educational managers, and agricultural extension systems to evaluate and improve the quality of training programs. This research is a quantitative study that, in addition to considering the frequency of expert opinions, used the weighted average method to aggregate expert opinions on Kirkpatrick level weights and effectiveness grading cut-offs for skill-based agricultural extension training courses.

Another notable finding is the homogeneity of opinions among agricultural extension and education specialists working in universities and AREEO. The Chi-

square test results showed that workplace had no significant effect on the choice of level weights or grading scale. This finding indicates a reasonable consensus among experts in this field in Iran, which increases the validity and generalizability of the results. In other words, regardless of organizational context, experts hold similar views on the relative importance of Kirkpatrick's levels. Iranian agricultural extension and education specialists clearly prefer a set of weights that balances behavior and results when evaluating training effectiveness. Unlike simply selecting the modal option, the weighted average method uses the full distribution of responses and thus takes minority viewpoints into account (Norman, 2010). Following the theory-driven empirical adjustment approach and the sensitivity testing framework (te Lintelo et al., 2014; Zardari et al., 2015), the weights derived from the weighted average were adjusted to theoretically aligned values while ensuring that the absolute deviation (8 percentage points, i.e., 4% of the maximum possible error) remained within an acceptable range and that the practicality of the composite index was enhanced. The sensitivity analysis was conducted to examine whether the proposed theoretical adjustment would materially influence the final effectiveness score. Such analyses are commonly recommended when empirical values are modified for theoretical or practical considerations (te Lintelo et al., 2014; Zardari et al., 2015). Therefore, to achieve full alignment with Kirkpatrick's theoretical foundations while maintaining proximity to the empirical findings, the following adjusted weights and final effectiveness formula are recommended for calculating the effectiveness of skill-based agricultural extension training courses (Table 8). These weights respect Kirkpatrick's theoretical principles, have acceptable expert support, and are practically applicable for evaluating agricultural skill-based courses.

The adjusted weights are 15%, 20%, 35%, and 30%, with behavior (application of knowledge on the farm) receiving the highest weight and reaction (satisfaction) the lowest. The formula assumes that the score for each level has been scaled or standardized (e.g., scores calculated on a 0–100 scale). This approach allows for the calculation of a total effectiveness score and objective comparison of different courses. Regarding the grading cut-off scale, Option C (with a threshold of 80 for "high effectiveness") received the highest level of agreement and frequency among experts, and the weighted average calculation also confirmed this classification. These cut-offs represent a balanced approach, neither as strict as Option A (90 for high) nor as lenient as Option D (75 for high), and are consistent with common grading systems in educational evaluation. The derived grading scale (Option C) is both the modal choice and the result of the weighted average, providing strong support for its use. The cut-offs (80, 60, and 40) are easy to understand, memorable, and compatible with conventional scoring systems. Therefore, Option C is recommended as the "effectiveness grading scale for skill-based agricultural extension training courses" (Table 9).

Table 6. Frequency distribution of effectiveness grading cut-off options

Option	High effectiveness	Moderate effectiveness	Weak effectiveness	Non-effective	Frequency	Percent
A	Score > 90	90 ≥ Score > 70	70 ≥ Score > 50	50 ≥ Score	2	3.9
B	Score > 85	85 ≥ Score > 70	70 ≥ Score > 55	55 ≥ Score	8	15.7
C	Score > 80	80 ≥ Score > 60	60 ≥ Score > 40	40 ≥ Score	22	43.1
D	Score > 75	75 ≥ Score > 55	55 ≥ Score > 35	35 ≥ Score	19	37.3
Total					51	100

Table 7. Weighted average calculation for determining effectiveness grading thresholds

Boundary	Product ($f_j \times t_j$)	n	Weighted average cut-offs	Rounded cut-offs
High–Moderate	4045	51	79.31	79
Moderate–Weak	3065	51	60.10	60
Weak–Non-effective	2085	51	40.88	41

Table 8. Weights for Kirkpatrick's effectiveness evaluation model levels for skill-based agricultural extension training courses

Level	Adjusted weight (%)
Reaction (R)	15
Learning (L)	20
Behavior (B)	35
Results (O)	30
Effectiveness formula:	**E** =(0.15 × R) + (0.20 × L) + (0.35 × B) + (0.30 × O)*

* R, L, B, O are the normalized scores (0–100) for reaction, learning, behavior, and results, respectively.

CONCLUSION

In conclusion, this study provides an evidence-based and locally calibrated framework for evaluating the effectiveness of skill-based agricultural extension training courses in Iran. The proposed composite effectiveness index assigns the highest weights to behavior and results, reflecting the primary goal of agricultural extension: sustainable on-farm change and improved productivity. Furthermore, the established grading cut-offs (high > 80, moderate 60–80, weak 40–60, and non-effective ≤ 40) provide a clear and interpretable scale for assessing training performance, while the consensus observed between university and research organization specialists confirms the robustness and generalizability of the proposed weighting and grading system. Consequently, adopting this standardized approach can enhance transparency, comparability, and evidence-based decision-making in agricultural training evaluation, ultimately contributing to more effective extension programs. To the best of the author's knowledge, this is the first study in Iran to propose an expert-derived weighting system and an effectiveness grading scale for Kirkpatrick's four-level evaluation model in agricultural extension training. The results of this study can serve as a "local standard" for evaluating the effectiveness of agricultural extension training courses in Iran, contributing to transparency, comparability, and evidence-based decision-making within the country's agricultural education and extension system.

Based on the findings, the following recommendations are provided in two categories: practical and research.

A) Practical recommendations (for policymakers, managers, and extension specialists)

Use of proposed weights in official evaluations: All institutions responsible for agricultural training and extension (AREEO, Agricultural Extension Coordination Offices, universities, and agricultural higher education centers) are advised to use the weights of 15%, 20%, 35%, and 30% for reaction, learning, behavior, and results, respectively, when evaluating the effectiveness of their training-extension courses. This will enable standardized comparison of different courses. The recommended weights have a theoretical basis, expert consensus, ease of understanding and application, and align with the ultimate logic and goals of agricultural extension.

Design of an integrated scoring system: It is recommended to develop software or an electronic system based on the composite effectiveness index (E = (0.15 × R) + (0.20 × L) + (0.35 × B) + (0.30 × O)) so that extension specialists can easily calculate and classify the effectiveness of delivered courses.

Prioritize behavior and results levels in course design: Given the high weights of behavior and results, course designers should focus most on components that lead to sustainable behavioral change on the farm and improvement in farmers' productive and economic performance. Simply generating satisfaction or theoretical knowledge is insufficient.

Use proposed cut-offs for standardized judgment: The cut-offs established in Table 9 should be used as a basis for managerial decision-making (e.g., continuing, revising, or terminating courses). Courses with an effectiveness score below 40 require fundamental revision.

Incorporate weights and cut-offs into evaluation guidelines: AREEO and the Deputy for Education and Extension of the Ministry of Agriculture can integrate these findings into national guidelines for evaluating extension courses.

B) Research recommendations (for researchers and graduate students)

Field validation of proposed weights: It is recommended to conduct an independent study using real data from dozens or hundreds of training-extension courses (with a large sample of participating farmers) and validate the proposed weights through data-driven statistical methods such as PCA or SEM.

Comparison of different weighting methods: Future research can compare the weights derived from the expert-based approach (the present method) with those from other methods such as AHP, Shannon entropy, and factor analysis to assess agreement and stability.

Investigate the effect of course characteristics on weights: It is recommended to examine whether course topic, duration (short-term vs. long-term), target group (smallholder vs. industrial farmers), and geographical region affect the weights of Kirkpatrick's levels.

Develop specific cut-offs for agricultural sub-sectors: The present study proposed general cut-offs for agricultural sciences. Subsequent studies can develop specific cut-offs for different sub-sectors (agronomy, horticulture, animal husbandry, fisheries, and natural resources).

International comparative study: It is recommended to conduct a similar study in other countries in the region to enable a regional model for weighting and effectiveness cut-offs in agricultural training.

Table 9. Effectiveness grading scale for skill-based agricultural extension training courses

Effectiveness score	Grade of effectiveness
Score > 80	High
60 < Score ≤ 80	Moderate
40 < Score ≤ 60	weak
Score ≤ 40	Non-effective/needs redesign

Examine the effect of contextual factors on expert opinion: Although workplace had no significant effect in this study, it is recommended to examine variables such as work experience, evaluation experience, gender, and experience in international events on experts' opinions.

Design and validation of standardized measurement instruments: Research can design and psychometrically validate standardized questionnaires for measuring each of the four levels (reaction, learning, behavior, and results) in various extension courses based on the proposed weights and cut-offs.

Structural relationship modeling among levels: Given recent methodological advances (Sathishwaran et al., 2025), it is suggested to examine causal relationships among Kirkpatrick's four levels in the context of Iranian agricultural education using SEM to extract direct and indirect weights more precisely.

This study has several limitations that should be acknowledged. First, the proposed weighting framework was developed based on the judgments of a relatively limited number of experts within the Iranian agricultural extension system. Therefore, the expert-derived weights and effectiveness rating scale should be interpreted within this specific institutional and national context and may not be directly generalizable to other sectors, organizations, or countries without further empirical validation. Second, the findings are inherently dependent on expert judgment, which, although appropriate for addressing the research objective, may be influenced by the experts' professional experiences and contextual perspectives. Furthermore, the questionnaire employed a limited number of structured response items designed to obtain independent expert judgments rather than detailed qualitative evaluations. Consequently, the proposed weights should be regarded as expert-derived recommendations that require further validation using larger and more diverse expert panels and advanced analytical methods.

FUNDING

This research was funded by Agricultural Research, Education and Extension Organization (AREEO), Shiraz, I. R. Iran.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization: Arman Bakhshi Jahromi; Methodology: Arman Bakhshi Jahromi; Software: Arman Bakhshi Jahromi; Validation: Arman Bakhshi Jahromi; Formal analysis: Arman Bakhshi Jahromi; Investigation: Arman Bakhshi Jahromi; Resources: Arman Bakhshi Jahromi; Data curation: Arman Bakhshi Jahromi; Writing—original draft preparation: Arman Bakhshi Jahromi; Writing—review and editing: Arman Bakhshi Jahromi; Visualization: Arman Bakhshi Jahromi; Supervision: Arman Bakhshi Jahromi; Project administration: Arman Bakhshi Jahromi; Funding acquisition: Arman Bakhshi Jahromi.

DECLARATION OF COMPETING INTEREST

The author declares no conflicts of interest.

ETHICAL STATEMENT

This research was conducted in accordance with the protocol and standards of research ethics, and the interviewees participated in the study with informed consent.

DATA AVAILABILITY

The data utilized in this study are outlined within the article.

ACKNOWLEDGMENTS

The author sincerely thanks all the respected professors and specialists in agricultural extension and education who participated in this research and provided their valuable opinions.

REFERENCES

Bakhshi Jahromi, A., & Heydari, M. (2022). Effectiveness of in-service training courses for employees of Fars Agricultural Jihad Organization. Proceedings of the 9th National Congress of Agricultural Extension, Natural Resources and Sustainable Environment, September 6-8, Tabriz, pp. 1-12. (In Persian)

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Long Grove, Illinois: Waveland Press.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36*(4), 1065-1105. https://doi.org/10.1177/0149206309352880

Cohen, R. J., & Swerdlik, M. E. (2021). *Psychological testing and assessment* (9th ed.). New York: McGraw-Hill.

Faisal-E-Alam, M., & Islam, A. R. M. T. (2025). Exploring the hierarchical framework of the Kirkpatrick model in training evaluation from a developing country. *Evaluation and Program Planning*, 113, 102696. https://doi.org/10.1016/j.evalprogplan.2025.102696

Hsu, C. -C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-8. https://doi.org/10.7275/pdz9-th90

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. Washington, DC: ATD Press.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.

Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). U. K: Pearson Education.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625-632. https://doi.org/10.1007/s10459-010-9222-y

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15-29. https://doi.org/10.1016/j.im.2003.11.002

Paul, S., Burman, R. R., & Singh, R. (2024). Training effectiveness evaluation: Advancing a Kirkpatrick model based composite framework. *Evaluation and Program Planning*, *107*, 102494. https://doi.org/10.1016/j.evalprogplan.2024.102494

Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational Assessment, Evaluation and Accountability*, *22*(3), 215-225. https://doi.org/10.1007/s11092-010-9098-7

Sathishwaran, R., Helen, S., Bonny, B. P., & Darsana, S. (2025). Structural equation modelling analysis of the Kirkpatrick training evaluation model. *The Journal of Agricultural Education and Extension*, 1-17. https://doi.org/10.1080/1389224X.2025.2573498

Taylor, M., & Bhasme, S. (2018). Model farmers, extension networks and the politics of agricultural knowledge transfer. *Journal of Rural Studies*, *64*, 1-10. 10.1016/j.jrurstud.2018.09.015

te Lintelo, D. J. H., Haddad, L. J., Leavy, J., & Lakshman, R. (2014). Measuring the commitment to reduce hunger: A hunger reduction commitment index. *Food Policy*, *44*, 115-128. https://doi.org/10.1016/j.foodpol.2013.11.005

Tiyawongsuwan, S., Lawthong, N., & Kanjanawasee, S. (2013). Standard setting for educational evaluation: concepts, methods and applications. *Kasetsart Journal of Social Sciences*, *34*(2), 225-240.

Zardari, N. H., Ahmed, K., Shirazi, S. M., & Yusop, Z. B. (2015). *Weighting methods and their effects on multi-criteria decision Making model outcomes in water resources management*. Springer.