Evaluation of a Science Teachers' Training Extension Program: Lesson Learning and Implications for Program Design

Emmanuel M. Preña
Cherrylyn P. Labayo
Bicol University Center for Policy Studies and Development, Legazpi City

Abstract

This study presented an approach to evaluating a higher education institution's teachers' training extension program (HEI) using the Kirkpatrick evaluation model. The adapted model provided an excellent framework for identifying the strengths and weaknesses of the training process. Findings revealed that the extension program was effective at the model's reaction level (Level 1), as evidenced by a high level of satisfaction. The Level 2 (learning criteria) and Level 3 (behavior) of the model were not successfully documented due to limitations of monitoring data of the extension program. However, the final results (Level 4) were examined using the Qualitative Impact Protocol (QUIP). According to the QUIP findings, the participation of teachers in various training and seminars on science and ICT topics was widely cited as a positive driver of change across the three domains of the training program. Most teachers made positive implicit statements that corresponded to the expected changes that the extension program aims to achieve, but they made no explicit reference to the project. The analysis provided the extension practitioners with a holistic understanding of the preparation, design, and implementation of similar future teachers' training extension programs in HEIs, focusing on the professional development of science teachers.

Keywords: Evaluation; extension; Kirkpatrick model; qualitative impact protocol; training program; science investigatory project, computer literacy

Introduction

Higher education institutions (HEIs) in the Philippines perform three interrelated functions: instruction, research, and extension. Most academic institutions' primary role is to teach. However, for research universities, research can be their primary function. Meanwhile, extension is the third function of HEIs that uses research findings to share new technologies and innovations with extension clients in a community (Quimbo & Sulabo, 2014). As producers of knowledge or hubs of innovations, HEIs are mandated to transfer knowledge or technology to improve the quality of the human life of Filipinos. The Commission on Higher Education [CHED] (2016) broadly defined extension as "the systematic transfer of technology, innovation or information generated by HEIs and its partners to seek solutions to specific developmental concerns" (p. 8). Thus, extension is the community engagement mission for all Filipino universities (Bernardo et al., 2012).

The modalities of delivery of extension programs by HEIs differ. The majority of these are training extension programs in education, health, social service, and livelihood. Like other community interventions, extension programs must be subjected to impact evaluation, which measures the outcomes attributable to an intervention. However, according to the literature, only a few publications look into extension programs' long-term benefits and social impact (Soska & Butterfield, 2013). Furthermore, conducting a reliable evaluation of an extension program presents several challenges. For example, Sermona and colleagues (2020) discovered that one of the challenges faced by Philippine HEIs in evaluating extension programs was the lack of a monitoring and evaluation design. As a result, the majority of the impact studies conducted on community extension programs were heavily reliant on beneficiary perception. Monitoring and evaluation are critical elements for evidence-based policymaking because they provide tools for verifying and improving the quality, efficiency, and effectiveness of policies and programs at various stages of implementation (Gertler et al., 2016).

The purpose of this study was to present a method for evaluating the Legazpi Port District II Extension Assistance Program for Science and Technology (LEAP for S&T). The Bicol University College of Science implemented this two-year training extension program from 2015 to 2017. The goal of this extension program was to provide support within the college's expertise to the capacity development needs of public science teachers. The program was pursued with the hope that by training teachers, they would become better educators. This will cascade effect on the students and hopefully on the educational institutions and community as a whole. In this light, the extension program focused on the capacity development of primary and secondary school teachers of Legazpi Port District II on (1) participation in science and technology (S&T) fairs, (2) quality of science investigatory projects (SIPs), and (3) computer literacy. Specifically, this study aimed to (1) assess the project's performance in terms of Kirkpatrick's Reaction, Learning, Behavior, and Results Model, (2) analyze differentiated narrative causal statements in the domains of science technology fair, science investigatory projects, and computer literacy; and (3) document lessons learned in evaluating extension efforts for teacher training programs on S&T and ICT to inform the design of future initiatives.

The outcomes of various evaluation studies of extension programs in the Philippine HEIs were usually measured using the conventional Likert scale (see Llenares, 2018; Montalbo, 2016; Salazar, 2020). However, in this study, the outcomes were assessed using the Qualitative Impact Protocol (QUIP), in which the narrative statements of beneficiaries served as evidence of attribution. Thus, this study also contributes to the literature on qualitative evaluation design.

Review of Related Studies and Literature

Teachers' quality is significant in improving the overall quality of education (Peterson, 2000). This demands interventions to improve the teaching skills of teachers so that educational objectives are met, particularly for public schools (Darling-Hammond, 1990; Darling-Hammond et al., 2017). Consequently, this translates to the performance of the students and the achievement of schools. The literature highlighted several interventions to improve the skills of teachers, one of which is the provision of teacher training (UNESCO, 2018a). Therefore, well-trained teachers are essential in achieving student learning outcomes (Seebruck, 2015; Sirait, 2016). Moreover, sustaining the teaching standards competence of teachers may require investment in training and development (Roberto & Madrigal, 2019).

In the Philippines, public schools encourage their students to conduct SIPs utilizing the scientific method through research as an entry to the annual S&T fair. This is organized by the Department of Education (DepEd) to promote S&T consciousness and a culture of innovation among the youth. This yearly event also identifies the most creative and innovative student researchers who will represent the Philippines in international competitions (DepEd Memorandum No. 134 s. 2018). Thus, this yearly event is a nationwide science research competition that begins at the school level and progresses to the division, regional, national, and international levels.

The SIP created by students with the help of their teachers is one of the requirements for a school to participate in the S&T fair. Autieri and colleagues (2016) described this S&T-related material as an instrument for students to make real-world connections and solve societal problems. Sanchez and Rosaroso (2019) documented the journey of secondary schools in SIP instruction through the lens of the teachers. Accordingly, science teachers must be competent to help their students develop practical projects using their critical thinking skills and "out of the box" perspective (see also Aparecio, 2018). Teachers frequently use their finances to conduct experiments and analyses in commercial laboratories because schools lack adequate laboratory facilities. Then, during the competition, science teachers instruct their students on how to disseminate the SIP results appropriately.

Meanwhile, ICT integration is also crucial in teaching. UNESCO (2018a) emphasized the importance of integrating ICT in schools and classrooms to transform pedagogy and empower students. It also suggests that students with high ICT literacy have higher academic achievement and that greater ICT literacy improves school learning outcomes (Lei et al., 2021). In science, ICT supports education in many ways. Databases, spreadsheets, and graphing tools can be used to teach science subjects (Demkanin et al., 2008). Thus, teachers have flexibility in teaching styles, allowing them to modify their material according to their students' needs and learning styles. Research evidence also pointed to increased motivation, interest, and attention span of students when learning is supported by ICT (University of York Science Education Group, 2002). Similar results were confirmed in employing ICT in physics teaching regarding acquisition of scientific concepts and the growth of scientific knowledge among students (Mohammed, 2013; Wu & Glaser, 2004).

Existing literature on teacher training evaluation indicates varied results. For example, Owston and colleagues (2008) found that science teacher training programs positively influenced teacher attitudes and content knowledge. Ertikanto and colleagues (2017) also asserted that elementary teachers acquired new skills from attending training programs. Teachers who have gained skills from training can help students perform better in their science investigation projects by mentoring them (Aparecio, 2018). However, the results reported no changes in science literacy and teachers' attitudes due to participating in similar training programs (Crall et al., 2013).

On the other hand, training programs in information and communication technology (ICT) positively impact teachers' attitudes and competence (Dela Fuente & Biñas, 2020; Karagiorgi & Charalambous, 2006). However, the competencies gained in ICT training did not significantly improve student learning and achievement (Karagiorgi & Charalambous, 2006). Although the objectives and contexts of teacher training programs vary in the literature, evaluation can be an effective tool for determining whether or not the intended goals were met and ensuring greater relevance to learners' work roles (Nemec, 2018).

The final stage of every training is an evaluation which measures the outcomes based on objectives. However, this aspect of the training process is frequently neglected (Giangreco et al., 2009). Most of the reasons in the literature why evaluation was being ignored were the lack of awareness of or access to methods and tools for the evaluation process (Mdhlalose, 2020; Eseryel, 2002). The failure of evaluation has also been linked to a lack of understanding of the requirement for a reliable evaluation (Berge, 2008) and purposely limiting evaluations to simple satisfaction metrics (Nemec, 2018). However, assessing training effectiveness does not need to be treated as a complex process (Praslova, 2010).

Recognizing the need for the teacher training program necessitates a credible evaluation to determine its effectiveness and future improvements. Therefore, an evaluation process can be successfully carried out using an appropriate methodology. Kirkpatrick's four-level model is considered the most widely used training evaluation framework that is straightforward, systematic, and practical (Kirkpatrick, 1976; Saad et al., 2013; Praslova, 2010). This model consists of four evaluation levels: Level 1-Reaction, Level 2-Learning, Level 3-Behaviour, and Level 4-Results.

Reaction refers to participants' perceptions of training after completing it (Kirkpatrick, 1976; Kirkpatrick & Kirkpatrick, 2006, 2009). This can be done by asking the participants to accomplish a post-training evaluation form. This level of evaluation does not yet measure what participants have learned but gauges the participants' interest, motivation, and attention levels (Smidt et al., 2009). The second level of evaluation is learning. This can be defined as "the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending the program" (Kirkpatrick & Kirkpatrick, 2006, p. 22). Thus, knowledge, skills, and attitudes are the three possible areas for framing the learning objectives of any program. This level may need a pretest and a post-test to examine whether or not learning has taken place. Assessing the participants' learning is a prerequisite for evaluating the next level of evaluation. However, this relationship can be erroneous. Arthur et al. (2003) attributed this idea that post-training environments may or may not provide opportunities for the learned material or skills to be demonstrated. The third level of evaluation is behavior. This pertains to the likelihood of transferring the knowledge and skills acquired when the participants return to their workplace. This level is quite challenging to assess compared to the first two levels. One critical consideration of this is when to conduct this level of evaluation. Two or three months after training is a good rule of thumb for some programs, but others may take up to six months or more to make the evaluation behavior more realistic (Axtell et al., 1997; Kirkpatrick and Kirkpatrick, 2006). In addition, level 3 does not need to be elaborate or scientific by simply asking a few people. The evaluator can use interviews or questionnaires, or both at this level. The fourth level, the most important and, at the same time, the most challenging part of the evaluation process, is the determination of final results that measure the impact the training has had. This might include improvement in, for example, the performance of an organization, quality of instruction, and reduced cost. The evaluation process at this level varies depending on the context of the project and the field where it belongs.

For over six decades, Kirkpatrick's model is still relevant in the training evaluation field. Kirkpatrick and Kirkpatrick (2006) asserted the applicability of the model whether the programs are conducted in education, business, or industry, regardless of the content of the programs and the type of participants. For academic institutions setup, in particular, Kirkpatrick and Kirkpatrick (2006) stressed that there is no attempt to change behavior. The purpose of the training and development is simply to increase knowledge, improve skills, and change attitudes. In these cases, only the first two levels apply. However, all four levels use if the purpose is to get better results by changing behavior. Consequently, most researchers have confirmed the need to measure the four criteria of evaluation to evaluate training outcomes accurately.

The current study attempted to explore the evaluation of a teacher training program as the core theme of an extension project of a higher education institution using Kirkpatrick's four-level model. The results criteria, particularly the most sought after by stakeholders, were measured using the Qualitative Impact Protocol (QUIP) outlined by Copestake and Remnant (2014). The QUIP addressed the limitations of establishing the counterfactual by asking the participants directly to generate reasonable evidence of causation. Thus, this paper contributes to the literature on impact evaluation by applying qualitative approaches in evaluating a short-term teacher training program.

Materials and Methods

Study Area

The study was implemented in Legazpi City, Albay. In particular, the areas covered were the participating elementary and secondary schools under the Legazpi Port District IIof the Department of Education (DepEd), Legazpi City Division. As mentioned in the introduction section, these schools were the target beneficiaries of the extension program.

Data Collection

The data used for this study were generated using documentary analysis and semi-structured interviews. The documents requested were reports about the progress and accomplishment of the extension program. The data for the survey was collected using a semi-structured interview. The questionnaire was patterned after the Qualitative Impact Protocol (QUIP) outlined by Copestake and Remnant (2017) from the University of Bath Centre for Development Studies. The questionnaire was composed of closed questions following the open-ended discussion. Open questions refer to generative questions asking the respondent to narrate any changes concerning the topic of interest (domains) from 2015 to 2019, if any changes have occurred, and the reasons for these changes. These domains were identified based on the extension project activities; for example, for the training on science investigatory projects, the outcomes of interest included participation in science technology fairs and quality of science investigatory projects of the school under evaluation.

Sensitivity and courtesy were observed before the interview. Respondents were provided informed consent emphasizing the anonymity and confidentiality of their responses. Furthermore, it was made clear that their participation was voluntary and that their information would remain confidential if they chose to end the survey at any time.

Sampling Design

The QUIP approach to sampling was through purposive sampling. Key Informant Interviews (KIIs) were changes and to whom or what they attributed these changes, which can be from multiple sources. Using the QUIP, research participants were allowed to share their experiences in an open, credible, and respectful way. There is no need for a control group since the evidence of attribution was sought through respondents' accounts of causal mechanisms linking the outcomes of interest to the training program (LEAP) alongside the other drivers of change rather than by relying on statistical inference based on participants' exposure to the project.

Analytical Design

The objectives of the study were analyzed using the four levels of evaluation introduced by Donald Kirkpatrick (1976). These four levels are reaction, learning, behavior, and results. The first three levels were evaluated by analyzing the available documents provided by the extension service providers. On the other hand, the results level was assessed using the QUIP. The data collected from QUIP was summarized using an Excel spreadsheet. Statements were coded into four types of statements: (1) Expl = change explicitly attributed to LEAP or explicitly named project activities; (2) Impl = change confirming or refuting the changes by which the LEAP aims to achieve, but with no explicit reference to LEAP or named project activities; (3) Inci = change attributed to other forces incidental to (not related to) the activities included in the LEAP's outcomes of interest; (4) Unat = change not attributed to any specific cause. The data was inductively summarized by identifying repetitions and patterns through immersion in the data. In addition, these statements were also coded according to whether respondents described their effects as positive or negative.

Limitations in Data

The study was initially designed to assess the impact of the extension program following Kirkpatrick's framework. However, during consultations with the concerned extension manager, evaluation was only conducted immediately after the training ended. There was no monitoring made sometime after the program ended. As of writing, the division handling extension programs in the university do not require monitoring reports after the project has been completed. Such limitations in data precluded the research team from conducting a full assessment. Nonetheless, gaps and insights highlighted in the paper provided opportunities for lesson learning. Results in the study can be used as a form of assessment to appraise the process of proposing, implementing, and evaluating extension projects not only at Bicol University but also in other HEIs.

Results and Discussions

Adaptation of Four-level Evaluation Model to Evaluate Training Extension Programs for Science Teachers

Of the four levels of Kirkpatrick's model, reaction and learning can be considered internal criteria as they focus on changes within the training program. These first two criteria can be quickly done, provided that guidelines for evaluating reaction and learning are appropriately designed. Behavioral and results, meanwhile, can be considered as external criteria as they focus on changes that occur after the program and can be influenced by factors other than learning, such as the organizational and economic contexts (Alliger et al., 1997; Kirkpatrick and Kirkpatrick, 2006; Praslova, 2010). Unlike the first two criteria, behavioral results are relatively difficult to evaluate as more time will be required to decide on its evaluation design. The implementation of the four levels must be done sequentially. Level 1 must be assessed first before doing level 2 and so on. By doing this, conclusions at every level will not be compromised.

Reaction and Learning Level

The first two levels of training evaluation reaction and learning are internal. The reaction criteria measure how participants react to the training they attended. Kirkpatrick and Kirkpatrick (2006) and other researchers described it as a measure of customer or client satisfaction. The assessment of client satisfaction involves the essential perspectives of "customers" about their experiences from attending the program. This level can measure one or multiple dimensions (Brown, 2007; Turner et al., 2018). For instance, in teacher training of an extension program for science teachers, the multidimensional constructs are the objectives of the activity, the usefulness/relevance of the activity, the contribution of the activity to community development, and the capability of the training provider. These indicators are helpful in the determination of trainees' satisfaction levels in which the results can be a basis for enhancing the quality of training programs (Mulder, 2001).

The trainees' reaction level was measured after attending each training using the self-administered questionnaires during the post-evaluation survey. Findings from this study showed that the trainees had a high level of satisfaction with each training they attended (Table 1). On a scale of 1 to 5, 1 being the lowest and 5 being the highest level of satisfaction, the average level of satisfaction of the trainees ranged from 4.62 to 4.95 from attending science investigatory project (SIP) and other science-related training. This indicates that science teachers had a positive reaction in the first level of Kirkpatrick's model regarding the specified multidimensional constructs of the training.

Many studies have highlighted the importance of client satisfaction in evaluation. Trotter (2008) found strong correlations between client satisfaction and a program's effectiveness. Hsieh and Guy (2008) and Morgan and Casper (2000) also asserted that clients rate higher levels of satisfaction when trainers are capable and comfortable performing the job. It appears that trainers played significant roles in the trainee's overall satisfaction perceptions of the training. The trainers of LEAP for S&T were faculty members of the Bicol University College of Science who served as extension service providers and were experts in biology, chemistry, physics, mathematics, and computer science. This corresponds to the finding of Bayar (2017) that the quality of trainers and training delivery affected teachers' satisfaction with mentoring activities. In addition, the relevance of the training was also found to be an indicator of teachers' professional growth and development (Lucas et al., 2017). The extension service providers provided the training according to the capacity development needs of science teachers in delivering substantial, motivating, and enhanced science courses to their pupils. For example, during the science investigatory project sessions, science teachers were given hands-on experience with the scientific method by being exposed to various biological, chemical, physical, and statistical procedures used to create science investigatory projects. Thus, the extension program was effective in attaining a positive response from teachers during the actual implementation of the series of training by considering the expertise of resource persons and facilitators who served as extension service providers.

Table 1

Summary of Training Evaluation Surveys on Teacher Training on Science-Related Topics

Training conducted	Date of Implementation	Number of Participants	Average Satisfaction
Science Investigatory Project 1	Sept. 4-6, 2014	42	4.95
Science Investigatory Project 2	Sept. 6-7, 2014	23	4.68
Training on Animal Handling	Feb. 9-10, 2015	18	4.62
Mosquito Identification	Feb. 9-10, 2015	38	4.64
Essential Microbiology using Household Tools	Feb. 9-10, 2015	18	4.65
Science Investigatory Project 3	Sept. 17-18, 2015	40	4.68
Research Writing Workshop	Sept. 17-18, 2015	42	4.95
Science Investigatory Project 4	Sept. 22-23, 2016	37	4.77

Source: Bicol University College of Science (2021)
Note: No data available for the computer literacy component

After assessing the trainees' satisfaction in the first level of Kirkpatrick's model, the next level must answer whether or not the trainees learned anything from attending the training program. The level of satisfaction alone does not necessarily mean the trainees acquired knowledge, developed their skills, and changed their attitudes. Thus, the learning criteria determine these expected outcomes from the participants after completion of the training. The expected learning outcomes were expounded further by Kraiger and colleagues (1993) and categorized into cognitive, skill-based, and attitudinal. The acquisition of knowledge falls under the mental aspect of learning outcomes, while the acquisition of technical skills is classified as a skill-based learning outcome. Finally, the attitudinal learning outcomes consist of factors such as participants' motivational disposition, self-efficacy, and goal setting that can be inferred as evidence of their development during training.

Potential training evaluation methods can be used to evaluate the learning outcomes of a training program, such as self-report measures, free recall measures, and pre and post-test measures – the most conventional and direct method. The pre and post-tests are typically used in higher education settings for assessing training effectiveness (Arthur et al., 2003). Although the procedures for evaluating learning outcomes are straightforward, the absence of these measures has direct implications in assessing Level 2 of Kirkpatrick's model. Therefore, a comparison of before and after training results that can indicate what changes have taken place will not be documented.

Evaluation at this level was not attained in the study. While pre- and post-tests are common forms of evaluating training programs to improve the participants' knowledge, the extension program providers failed to provide assessment tools in the form of pre- and post-evaluation that could have assessed the teachers' knowledge, skills, and attitudes. This suggests that improvements should be made in designing future training programs by including pre- and post-test in extension proposals.

Behavioral and Results Level

Applying the other two criteria (behavior and results) to teachers' training as an extension program requires adapting the model to a higher education institution's specific context and purposes (Praslova, 2010). Behavioral and results criteria are external and thus are influenced by various factors other than learning. Among the criteria of the model, the data for these two are the most difficult to obtain, and such data are rarely completed. The behavioral criteria, in particular, identify the effects of training on the work performance of the participants in their workplace. This can be analyzed using quantitative and qualitative methods, especially in the education sector (Kirkpatrick & Kirkpatrick, 2006). In the case of science teachers in elementary and secondary schools, one potential evidence to evaluate behavioral materials and grading student works are some indicators to evaluate the behavioral criteria. Hence, follow-up data must be secured before evaluation. Typically, the best time to measure the behavioral change in participants, according to Kirkpatrick and Kirkpatrick (2006), is at least three months after the training. In a public school setting, this can be a whole school year comprising nine to 10 months in which S&T fairs are held in succession at different levels and periods – school level (September), division (October), regional (November), and national (February) – DepEd Memorandum No. 134 s. 2018.

This study was based on a two-year extension program focused on teachers' training that ended in 2017. Another constraint in this study was the unavailability of a post-training evaluation or follow-up data from the participants. Thus, the transfer of knowledge from the participants was not documented. According to the university's extension director (R. Zoilo, personal communication, October 13, 2021), follow-up data within a prescribed period after training is not standardized yet in the monitoring and evaluation system. Consequently, assessing the behavioral outcomes of teachers at the time of evaluation cannot be realized for three reasons: (1) depending on qualifications, public school teachers in elementary and secondary can be transferred to other stations, (2) some teachers who attended the training already retired from service, and (3) some students they taught already graduated. It was already five years since the project ended. Hence most of them were difficult to trace. As with the learning criteria, findings at this level suggest improvements in the implementation of extension programs in the university by requiring evaluation reports after the extension program has ended for some time.

The last and final level of the model is the final results that were accomplished because of the training program. This level is the outcome of the training program that usually focuses on organizational contexts. Therefore, it is necessary to consider the expected outcomes from the stakeholders' perspectives. Although results criteria in education may include a wide range of outcomes that expected outcomes from the stakeholders' perspectives. Although results criteria in education may include a wide range of outcomes that benefit the individual and the society, it is still imperative to be more specific. One way to determine the desired outcomes is to revisit the objectives of the training program. Hence, the evaluation of this study had four involved parties benefited: the elementary and secondary science teachers as the direct beneficiaries who attended the training program, their immediate students as indirect beneficiaries through cascading, the schools as their workplace, and the community where they live. These parties were the potential subjects for level 4 evaluation.

Kirkpatrick and Kirkpatrick (2006) suggested one guideline for evaluating results if the proof is not possible – be satisfied with the available evidence. The difficulties in assessing level 4 of the model were evident in the literature (see Berge, 2008; Mohamed et al., 2012). Some training programs only evaluated the reaction (Level 1), sometimes up to a learning (Level 2), assuming that it was already sufficient to assess whether or not the trainees had positive reactions to the training program. It could be assumed that trainees learned from the training, improved their work performance, and positively contributed to organizational results (Reio et al., 2017). These assumptions were turned down, considering there was no guarantee that a positive reaction assured learning, favorable behavioral outcomes, and better organizational results (Kirkpatrick & Kirkpatrick, 2006).

Among the potential parties involved in the teachers' training program, the teachers handling science subjects for at least five years in the participating schools were selected for level 4 evaluation. In this study, the outcome of the training program was evaluated through the key informants' perceptions of the changes from 2015 to 2019 in participating schools (Table 2). The year 2020 was intentionally excluded because of the mobility restrictions imposed by the authorities in light of the COVID-19 pandemic. The data obtained from the respondents using the QUIP indicated a mixed picture of changes except for computer literacy. Six respondents reported that there were positive changes observed in the participation of schools in S&T fairs. Four also noted that the quality of their science investigatory projects made by their students significantly improved. This confirms the findings of Sanchez and Rosaroso (2019) that science teachers were seen to be instrumental in the SIP process as they instill basic research skills in students. Also evident in this study was the assistance and mentorship to pupils in creating a SIP that used minimal resources in low-resource schools. Meanwhile, positive changes in the use of ICT technologies in in struction were attributed to most of the respondents with computer literacy training. The literature also confirmed the positive effects of ICT training on teachers' confidence and teaching effectiveness (Galanouli et al., 2004; Xu & Chen, 2016).

Table 2
Responses to Closed Questions in QUIP

Responses	Participation in Science Technology Fair n=23	Quality of Science Investigatory Projects n=23	Computer Literacy n=23
Better	6	4	21
No change	6	6	0
Worse	1	1	0
Not sure	5	6	0
No response	5	6	2

Source: Author's calculations

The self-reported attribution, in the form of narrative statements from respondents, was explored to determine the outcome of the extension project. The narrative statements for each training component were inductively grouped and then systematically tabulated the drivers of change mentioned by at least two respondents. The main drivers identified were based on the three domains of the training program: (1) participation in S&T fairs, (2) quality of SIPs, and (3) computer literacy (Table 3). Participation in training and seminars was widely cited as a positive driver of change in the threedomains. The respondents viewed that their participation in various training and seminars was effective concerning improved participation in S&T fairs. This also made them more efficient in preparing learning materials for their learners. Limited resources, from financial to human, including time, were still identified as a constraint why some schools opted not to participate in any S&T fairs. However, it is important to note that the magnitude of the outcome among the beneficiaries in this study remains unknown. Therefore, the QUIP should be viewed as a method of contribution analysis rather than impact assessment (Copestake & Remnant, 2014)

Table 3
Most Widely Cited Positive and Negative Drivers of Change

Domain	Positive	Negative
Participation in science and technology fairs	Participation in training and seminars (4) Access to the internet and gadgets (3)	Limited resourc-es (2)
Quality of science Tech-nology projects	Participation in training and seminars (4)	Limited resourc-es (4)
Computer literacy	Participation in training and seminars (19) Self-learning (4)	

Source: Author's calculations

Causal Statements of Respondents from KII

The narrative statements gathered from the KIIs were extracted to generate the number and the type of cause-and-effect statements that served as evidence of attribution (Table 4). Many respondents volunteered statements about positive drivers of change and most of which were implied. These implied positive drivers of change confirmed the outcomes of interest the teachers' training program aims to achieve, but with no explicit reference to the project or named project activities. Most respondents explained how the training programs improved their professional development, student achievement, and teaching strategies. For example, one respondent believed that "training and seminars help improve their continuing professional development and learners' academic performance, especially in mathematics and science". One respondent also felt that her "students became more motivated and interested to learn and discover new things." On their participation in S&T fairs, one respondent stated that "the school has improved its participation by regularly joining to S&T fairs because of their students' output as an entry to the event." The quality of SIPs was also enhanced because of students' exposure to several S&T fairs and assistance from their teachers. For instance, one respondent reported that "the scientific inquiry skills of pupils were developed. This encouraged them to think and find solutions in creative ways and come up with an excellent SIP." This was supported by another respondent saying that "pupils developed the sense of confidence, responsibility, and trust among teammates." Another advantage of taking part in this teacher training program was the ability to integrate ICT into their classroom arrangements. For example, one respondent asserted that "through ICT training and seminars they attended, they were able to discover useful computer applications that helped them in making their teaching styles more engaging to learners."

Even though most of the respondents' positive narrative statements made no direct reference to the extension program, positive changes occurred in the three domains that the extension program aspired to achieve. The results from attending training programs in this study are consistent with the literature, which emphasizes the importance of teachers' professional development through training and the positive influence of mentorship on learners (Arnesson & Albinsson, 2017). Cribbs and colleagues (2020) asserted the benefits of attending a teacher training program in increasing inquiry-based instruction in mathematics and science classrooms. It aligns with Darling-Hammond and colleagues (2017) that effective professional development is critical for accomplishing student achievement goals. Similarly, computer and internet literacy integration in classrooms can generate a strong positive attitude and appreciation among learners (Nuncio et al., 2020).

Documentation of Lessons Learned in Evaluating Training Extension Programs

This study met several constraints at all levels of the research that must be documented to come up with a lesson-learned framework design. This framework identifies areas where improve ments can be made in implementing similar future training extension programs. Thus, this section summarizes the lessons learned from the teachers' training program implemented by the Bicol University College of Science that must be considered. The lessons learned were categorized into four factors that resulted in the sub-optimal achievement of the impact of the teachers' training program (Figure 1). These factors constitute the project design,

Table 4
Frequency of Causal Statements of Respondents from KII

Indicators	Positive				Negative			
	Explore	Impl	Inci	Unat	Explore	Impl	Inci	Unat
Participation in Science & Technology Fair	0	6	2	1	2	0	1	3
Quality of Science Investigatory Projects	1	2	1	0	0	0	0	4
Computer Literacy	0	20	2	0	0	0	0	0

Source: Author's calculations

implementers, the teachers and students, and the confounding factors affecting the causal links between the project and its intended impact indicators. The analysis was based on the United Nations Development Programme (UNDP) Evaluation Guidelines (IEO UNDP, 2021).

Project design factors. The project design failed to capture the critical components of the evaluation process in the proposal stage of LEAP for S&T. This is because the proposal template for the extension project does not require an impact evaluation design. Thus, recognizing the importance of conducting an impact evaluation after completing a training extension program necessitates a well-defined impact evaluation design before its proposed commencement. An established evaluation design with a clear program results framework, SMART performance indicators, an impact evaluation method, and the timings and schedules for each stage, such as monitoring and follow-up data, can avoid constraints the LEAP for S&T experienced during its evaluation.

Project implementers factors. The roles of the project implementers in every step of the training management cycle are important in achieving the shared goal by meeting the project's objectives. The training management cycle comprised three significant steps: Step 1: Planning; Step 2: Implementation; and Step 3: Evaluation (JICA, n.d.). The first two steps of the project were carried out accordingly. However, the evaluation step was found to be challeng ing. The project implementers missed collecting the monitoring data necessary for evaluation, particularly for Kirkpatrick's model's learning and behavior level. This can be corrected when designing the project at the proposal level.

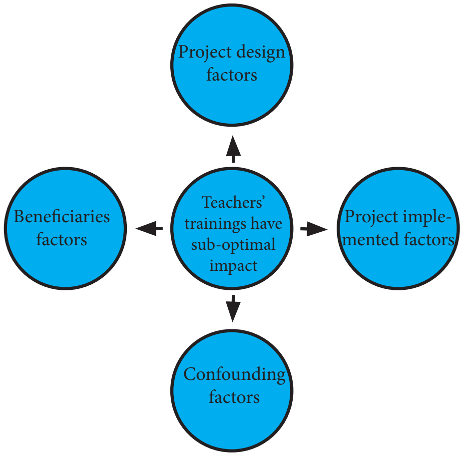
Beneficiaries factors. The beneficiaries of the training program include the teachers and students. Some of the teachers, including the school principals who directly benefited from attending the teachers' training program, were already transferred to other schools and retired from service during visits to schools for the evaluation. Thus, it was hardly impossible to locate the teachers in this case for them to participate in the current evaluation study. Similarly, some students benefited from the SIP component of LEAP for S&T through cascading of teachers who had already graduated. Also, it was not feasible to contact the students in this case.

Confounding factors. One main challenge encountered during the evaluation was the mobility restrictions because of the COVID-19 pandemic. Given this extreme situation, the evaluation was redesigned by excluding the respondents' perceived changes in 2020 to avoid attributions concerning the pandemic. Confounding factors also include similar interventions provided by institutions other than the project. Given its nature and contexts, this lesson-learned framework is not conclusive to other teachers' training programs.

However, this can be applied and enhanced for future similar extension projects regardless of the type of beneficiaries.

Figure 1

The Framework of Lessons Learned for LEAP and S&T:



Conclusions and Recommendations

This study illustrated that the Kirkpatrick evaluation model provided an excellent framework for identifying the strengths and weaknesses of the training process for science teachers in public schools. Based on the findings of this study using the adapted Kirkpatrick model, the LEAP for S&T was effective at Level 1 as it gained positive reactions in terms of the specified multidimensional constructs of the training. The teachers had a high level of satisfaction with each training they attended. However, the level of satisfaction does not necessarily equate to the acquisition of knowledge, development of skills, changes in attitudes of the teachers, as well as the transfer of knowledge to workplaces. With the absence of pre and post-tests for framing the learning objectives as well as the follow-up data for measuring the behavioral outcomes, the learning (Level 2) and behavior criteria (Level 3) were not successfully documented before the implementation of the last and final stage of evaluation (Level 4). However, these gaps and insights provided opportunities for lesson learning and implications for the program design of extension activities in HEIs.

Results from QUIP revealed that the participation in various training and seminars on science-related and ICT-related topics of teachers was widely cited as a positive driver of change among the three domains of the training program: schools' participation in S&T fairs, the quality of SIPs, and computer literacy. Several respondents provided positive implicit statements conforming to the expected changes the LEAP for S&T aimed to achieve, but with no explicit reference to the project or named project activities. However, the magnitude of the impact contribution of the training program was still uncertain. Thus, the narrative accounts of drivers of change can be viewed as evidence of attribution or a method for contribution analysis rather than impact assessment (Copestake & Remnant, 2014).

The lesson-learned framework indicated constraining factors that resulted in the sub-optimal achievement of the impact of the teachers' training program. These factors constituted how the project design was formulated, how well-known the project imple menters were in the evaluation process, the situation of the beneficiaries, and the confounding factors affecting the causal links between the project and its intended impact indicators. Extreme events affecting the data collection were also considered in meeting the desired results. The identified constraining factors provided a holistic understanding for extension practitioners in preparing, designing, and implementing extension programs for science teachers.

The evaluation offered a space for improvements in implementing future extension programs by considering backward design in planning and determining the desired outcomes. In this manner of planning, extension practitioners will be informed of what data should be collected before, during, and after the completion of an extension program. This will ensure the reliability of the results of an evaluation by appropriately defining the program indicators. In future research, a more robust analysis of measuring the impact of a training program can be done by implementing both the qualitative and quantitative approaches of impact evaluation. These two approaches will provide a more profound and broader understanding of the effects of training programs.

References

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, *50*(2), 341–358. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x

Aparecio, M.B. (2018). Mentoring, self-efficacy, and performance in conducting investigatory projects: a mixed-method analysis. *Asia Pacific Institute of Advanced Research*, *4*(2), 65-76. DOI: 10.25275/apjcectv4i2edu7

Arnesson, K. & Albinsson, G. (2017). Mentorship – a pedagogical method for integration of theory and practice in higher education. *Nordic Journal of Studies in Educational Policy*, *3*(3), 202–217. doi:10.1080/20020317.2017.1379346

Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: the relationship between reaction and learning criteria. *Educational Psychology*, *23*, 275–285.

Autieri, SM., Amirshokoohi, A. & Kazempour, M. (2016). The science-technology-society framework for achieving scientific literacy: An overview of the existing literature. *European Journal of Science and Mathematics Education*, *4*(1), 75-89.

Axtell, C.M., Maitlis, S., & Yearta, S.K. (1997). Predicting immediate and longer-term transfer of training. *Personnel Review*, *26*(3), 201–213.

Bayar, A. (2014). Satisfaction of Teachers with Mentoring Programs in Turkey. *The Anthropologist, 18*(2), 299–308. doi:10.1080/09720073.2014.11891547

Bernardo, M.A.C., Butcher, J., & Howard, P. (2012). *An international comparison of community engagement in higher education*, *32*(1), 187–192. doi:10.1016/j.ijedudev.2011.04.008

Berge, Z.L. (2008). Why it is so hard to evaluate training in the workplace. *Industrial and Commercial Training*, *40*(7), 390–395. doi:10.1108/00197850810912270

Brown, K. (2007). Developing, Training and evaluating employees. In Encyclopedia of Industrial and Organizational *Psychology;* Rogelberg, S.G., Ed.; SAGE Publications Inc.: Thousand Oaks, CA, US, 820–823.

CHED. (2016, October 3). CHED Memorandum Order 52, Series 2016: Pathways to Equity, Relevance and Advancement in Research, Innovation and Extension in Philippine Higher Education.

Copestake, J., & Remnant, F. (2017). *Qualitative Impact Assessment Protocol (QUIP)*. https://bathsdr.org/wp-content/uploads/2017/09/Revised-QUIP-briefing-paper-July-2017.pdf

Copestake, J., & Remnant, F. (2014). Assessing rural transformations: Piloting a qualitative impact protocol in Malawi and Ethiopia. Bath Papers in International Development and Wellbeing, No. 35, University of Bath, Centre for Development Studies (CDS), Bath.

Crall, A. W., Jordan, R., Holfelder, K., Newman, G. J., Graham, J., & Waller, D.M. (2013). The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. *Public Understanding of Science, 22*(6), 745–764. doi:10.1177/0963662511434894

Cribbs, J. D., Day, M., Duffin, L., & Cowley, K. (2020). Mathematics and science teacher training program improves inquiry practices for teachers. *School Science and Mathematics, 120*(8), 443–455. doi:10.1111/ssm.12437

Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teach ers* (pp. 17–32). Sage.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.

Dela Fuente, J. A. & Biñas, L. C. (2020). Teachers' competence in information and communications technology (ICT) as an educational tool in teaching: An empiricalanalysis for program intervention. *Journal of Research in Education, Science and Technology, 5*(2), 61-76.

Demkanin, P., Kibble, B., Lavonen, J., Guitart, J., & Turlo, J. (2008). *Effective use of ICT in Science Education*. School of Education University of Edinburgh. http://fizyka.umk.pl/~pdf/EU_ISE/files/new/EUISEBookHR.pdf

DepEd Memorandum No. 134 s. 2018. National Science and Technology Fair for School Year 2018-2019. https://www.deped.gov.ph/wp-content/uploads/2018/08/DM_s2018_134.pdf

Ertikanto, C., Herpratiwi, Yunarti, T., & Saputra, A. (2017). Development and evaluation of a model-supported scientific inquiry training program for elementary teachers in Indonesia. *International Journal of Instruction, 10*(3), 93-108. https://doi.org/10.12973/iji.2017.1037a

Eseryel, D. (2002). Approaches to evaluation of training: Theory & Practice. *Journal of Educational Technology & Society*, *5*(2), 93–98. http://www.jstor.org/stable/jeductechsoci.5.2.93

Galanouli, D., Murphy, C., & Gardner, J. (2004). Teachers' perceptions of the effectiveness of ICT-competence training. *Computers & Education*, *43*(1-2), 0–79. doi:10.1016/j.compedu.2003.12.005

Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., & Vermeersch, C.M.J. (2016). *Impact evaluation in practice* (2nd ed.).Washington, DC: Inter-American Development Bank and World Bank. doi:10.1596/978-1-4648-0779-4.

Giangreco, A., Sebastiano, A., Peccei, R., Giangreco, A., Sebastiano, A., & Peccei, R. (2009). Trainees' reactions to training: An analysis of the factors affecting overall satisfaction with training. *The International Journal of Human Resource Management*, *20*(1), 96–111. doi: 10.1080/09585190802528417

Hsieh, C.W., Guy, M.E. (2008). Performance outcomes: The relationship between managing the "Heart" and managing client satisfaction. *Review of Public Personnel Administration*, *29*(1), 41–57. doi:10.1177/0734371X08326767

IEO UNDP. (2021). *UNDP evaluation guidelines.*

JICA. (n. d.). *Manual on Training Evaluation.*

Karagiorgi, Y., & Charalambous, K. (2006). ICT in-service training and school practices: In search for the impact. *Journal of Education for Teaching, 32*(4), 395-411, doi: 10.1080/02607470600981995

Kirkpatrick, D.L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301-319). McGraw-Hill.

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

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2009) Evaluating: Part of a ten-step process. *In Evaluating training programs*. Berrett-Koehler Publishers.

Kraiger, K., Ford, J.K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *J. Appl. Psychol, 78*, 311–328.

Lei, H., Xiong, Y., Chiu, M. M., Zhang, J., & Cai, Z. (2021). The relationship between ICT literacy and academic achievement among students: A meta-analysis. *Children and Youth Services Review*, *127*, 106123. doi:10.1016/j.childyouth.2021.106123

Llenares, I. I., & Deocaris, C. C. (2018). Measuring the impact of an academe community extension program in the Philippines. *Malaysian Journal of Learning and Instruction*, *15*(1), 35-55.

Lucas, R.I., Promentilla, M.A., Ubando, A., Tan, R.G., Aviso, & K., Yu, K.D. (2017). An AHP-based evaluation method for teacher training workshop on information and communication technology. *Evaluation and Program Planning*, *63,* 93–100. doi:10.1016/j.evalprogplan.2017.04.002

Mdhlalose, D. (2020) An evaluation of the impact of training and development on organisational performance: A case study of the Gauteng Provincial Department of Economic Development. *Journal of Human Resource and Sustainability Studies*, *8*, 48-74. doi: 10.4236/jhrss.2020.81004.

Mohamed, R. & Alias, A.A.S. (2012). Evaluating the effectiveness of a training program using the four level Kirkpatrick model in the banking sector in Malaysia. *Proceedings of the 3rd International Conference on Business and Economic Research, Bandung, Indonesia*, 1–15.

Mohammed, S. A. (2013). Promoting scientific literacy by using ICT in Science Teaching. *International Education Studies*, *6* (9), 175-186. doi:10.5539/ies.v6n9p175

Montalbo, E. (2016). Impact assessment of the community extension programs of AB Mass Communication and Paralegal Studies towards community development. *IJABER*, *14*(5), 3397-3441.

Morgan, R.B., & Casper, W.J. (2000). Examining the factor structure of participant reactions to training: A multidimensional approach. *Hum. Resour. Dev*. Q, *11*, 301–317.

Mulder, M. (2001). Customer satisfaction with training programs. *Journal of European Industrial Training, 25*(6), 321–331. doi:10.1108/03090590110401791

Nemec, P.B. (2018). Training evaluation. Psychiatric Rehabilitation Journal, *41*(1), 80–81. https://doi.org/10.1037/prj0000291

Nuncio, R. V., Arcinas, M. M., Lucas, R. I. G., Alontaga, J. V. Q., Neri, S. G. T., & Carpena, J. M. (2020). An e-learning outreach program for public schools: Findings and lessons learned based on a pilot program in Makati City and Cabuyao City, Laguna, Philippines. *Evaluation and Program Planning*, 101846–. doi:10.1016/j.evalprogplan.2020.101846

Owston, R.D., Sinclair, M., & Wideman, H. (2008). Blended learning for professional development: An evaluation of a program for Middle School Mathematics and Science teachers. *Teachers College Record*, *110*(5), 1033-1064.

Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Corwin Press.

Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education. *Educ Asse Eval Acc*, *22*, 215–225. https://doi.org/10.1007/s11092-010-9098-7

Quimbo, M.A. T., & Sulabo, E.C. (2014). Research productivity and its policy implications in higher education institutions. *Studies in Higher Education. International Journal of Educa tional Development*, *39*(10), 1955–1971. doi:10.1080/030750 79.2013.818639

Reio, T.G., Rocco, T.S., Smith, D.H., & Chang, E. (2017). A critique of Kirkpatrick's Evaluation Model. *New Horizons in Adult Education and Human Resource Development, 29*(2), 35–53. doi:10.1002/nha3.20178

Roberto, J., & Madrigal, D. (2018). Teacher quality in the light of the Philippine professional standards for teachers. *Philippine Social Science Journal*, *1*(1), 67-80. https://doi.org/10.52006/main.v1i1.13

Saad, A., & Mat, N.B. (2013). Evaluation of effectiveness of training and development: The Kirkpatrick model. *Asian J. Bus. Manag. Sci., 2*, 14–24.

Salazar, T. B. (2020). An impact study of the community extension programs in a state college in the Philippines. *International Journal of Educational Sciences*, *29*(1-3), 16-23.

Sanchez, J.M.P. & Rosaroso, R.C. (2019). Science Investigatory Project Instruction: The Secondary Schools' Journey. The Normal Lights, *13*(1), 56 – 82

Seebruck, R. (2015). Teacher quality and student achievement: A multilevel analysis of teacher credentialization and student test scores in California High Schools. *McGill Sociological Review*, *5*, 1-18.

Sermona, N.L.D., Talili, I.N., Enguito, R.C., & Salvador, M.F. (2020). Implementation of extension services in select state universities and colleges in the Philippines. *Sci.Int.(Lahore)*, 32(6), 609-614.

Sirait, S. (2016). Does teacher quality affect student achievement? An empirical study in Indonesia. *Journal of Education and Practice, 7*(27).

Smidt, A., Balandin, S., Sigafoos, J., & Reed, V.A. (2009). The Kirk patrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability*, *34*(3), 266–274. doi:10.1080/13668250903093125

Soska, T., & Butterfield, A. K. J. (2013). *University-community partnerships: Universities in civic engagement*. Routledge.

Trotter, C. (2008). What does client satisfaction tell us about effectiveness?. *Child Abuse Review*, *17*(4), 262–274. doi:10.1002/car.1038

Turner, M.R., Watts, L.L., Steele, L.M., Mulhearn, T.J., Torrence, B.S., Todd, E.M., Mumford, M., & Connelly, S. (2018). How did you like this course? The advantages and limitations of reaction criteria in ethics education. *Ethics & Behavior, 28*(6), 483–496.

UNESCO. (2018a). *World Teachers' Day: Quality education requires well-trained teachers*. https://iite.unesco.org/news/world-teachers-day-quality-education-re quires-well-trained-teachers/

UNESCO. (2018b). *UNESCO ICT Competency Framework for Teachers*. https://unesdoc.unesco.org/ark:/48223/pf0000265721

University of York Science Education Group. (2002). *ICT in support of Science Education: A practical user's guide*. York Publishing Services Ltd. https://www.york.ac.uk/org/seg/about_us/about_us_images/ICTinSupport.pdf

Wu, Z. and Glaser, R. E. (2004). Software for the synergistic integration of science with ICT education. *Journal of Information Technology Education, 3,* 325–339.

Xu, A., & Chen, G. (2016). A study on the effects of teachers' information literacy on information technology integrated instruction and teaching effectiveness. *Eurasia Journal of Mathematics. Science and Technology Education, 12*(2), 335-346. https://doi.org/10.12973/eurasia.2016.1222a