Evaluation of Teacher Professional Learning Workshops on the Use of Technology - A Systematic Review

Alireza Ahadi, Matt Bower, Jennifer Lai, Abhay Singh, Michael Garrett

January 24, 2022

Abstract

Teacher professional learning workshops have been frequently used to prepare in-service and pre-service teachers for effective use of technology in education. Evaluation of these workshops is crucial to identify the effectiveness of these programs in terms of improving teaching skills, increasing knowledge, changing attitudes, and developing capabilities which support the achievement of student learning outcomes. Multiple approaches for evaluating professional development programs have been developed, though each with different emphases and theoretical positioning. In this systematic review, 41 research-oriented teacher professional development workshops on technology use were critically analysed to understand how such workshops are evaluated. This study examines content evaluation, instrument types, and common professional development frameworks in teacher professional development workshops and reports on their usage and alignment. Based on the findings, the majority of papers in the systematic review did not deploy comprehensive professional development evaluation models to assess teacher professional development workshops. The majority of studies did not report on the use of established instruments for the purposes of data collection. It was further observed that the majority of these studies did not attempt to evaluate different dimensions of teacher change or different dimensions related to evaluation.

1 Introduction

Teacher professional learning is critical to the development of the education field, because it makes teachers more contemplative and aware of their on-going learning. This is especially common and true in the rapidly evolving area of technology use in education, since such programs improve teachers' ability to engage with students in a more interactive fashion and provide students with quality learning environments (Bruce 2003). In particular, research in the educational technology field has focused on how teachers can seamlessly integrate technology into the classroom in effective and meaningful ways (Kostiainen et al.

2018). Different teacher professional learning models have been introduced and implemented extensively in the past, such as Coaching and Mentoring, Face-to-Face training, Train-the-Trainer model and web-based professional development models, all of which are commonly used in designing teacher professional learning workshops (Poplin 2003).

Evaluation of teacher professional development workshop is a crucial step in the development of such programs. The evaluation results not only help program organisers to make improvements and maintain the high quality of their programs but also reveals whether or not the approaches being adopted are yielding the intended results. However there is a great variety of ways that teacher technology workshops can be evaluated, and in many cases the approach to evaluation is ad-hoc and un-theorised (see review below). In the era of COVID-19 it is even more important that online professional learning is rigorously evaluated to inform and improve future practice and offerings, as more professional learning that may have traditionally been offered face-to-face is now being conducted in online mode. This paper conducts a systematic review of the various approaches to evaluation that have been used within the professional development research literature with respect to teacher technology workshops. This provides educators and researchers with a more accurate sense of workshop evaluation practices so that they may conduct more effective professional learning workshop evaluation. The overarching research question for this study is "How are professional development workshops relating to the use of technology in teaching evaluated?". In particular, this paper reports a systematic review of teacher professional learning technology workshops to address the following research questions:

RQ1: What are the constructs examined when evaluating teacher professional learning technology workshops?

RQ2: What are the professional learning frameworks used in studies that evaluate educational technology workshops?

RQ3: What are the prevailing instruments that are used to evaluate educational technology workshops?

The findings from this study can help teacher educators and workshop developers optimise their workshops by adopting more robust and evidence based professional learning evaluation.

2 Motivation for this study

Fundamentally, teacher professional learning programs aim to affect change. Educators attending a professional development workshop, an educational seminar or a continuing professional development program are seeking positive change,

just as the facilitators who are conducting the workshops aim to affect change in their participants. Clarke and Hollingsworth (2002) characterise the concept of teacher change as being a result of training, adaptation, personal development, local form, systemic restructuring, or learning. The influence of change theory on teacher professional development design, implementation and evaluation has been significant. Guskey's seminal linear path model of teacher change (Guskey 2002*a*,*b*), is founded on teacher change theory. Guskey states that changes in teacher's attitudes and beliefs can be the result of positive student outcomes, which in turn can be the consequence of completing a professional learning program that impacts teacher's classroom practice. More broadly, from an educational point of view, these changes can relate to perceptions, knowledge, skills, attitudes, or a variety of other measurable variables. Hence, program evaluation should focus on revealing the change that occurs to these different parameters.

The judgement of the value in the observed change is the fundamental core of program evaluation. Accordingly, educational program evaluation seeks to reveal the value and worth of a program with respect to the observed change (Cook 2010). From a teacher change stand point, there are certain characteristics that enable or potentially inhibit teachers from effectively leveraging technology in their classes. Impediments to successful technology integration include "first-order" barriers such as access to resources, training and support, but also internal or "second-order" barriers such as teacher confidence, perceived value of technology and perceptions about how students learn (Ertmer et al. 2012). Hence, the use of appropriate teacher professional development evaluation models which take these teacher areas into account is fundamental.

3 Background

The design of formalised professional development evaluation methods can be traced back to 1959 when Kirkpatrick introduced his seminal four level evaluation framework (Kirkpatrick & Kirkpatrick 2006). The model evaluates a teacher professional development program on four different levels: (1) reaction (learner satisfaction with the program); (2) participant's learning (for instance increased knowledge, skills obtained, attitudes changed); (3) changes in learner behavior; and (4) the program's impact on business objectives. This model has roots in reductionism in the way that its underlying assumption is the causality relationship between the professional development program and its outcomes, and as a result it has been criticised by professional development evaluation theorists for not considering other parameters which can possibly impact the final program's outcome (Holton III 1996) (for instance, differences in prior knowledge before attempting the program, different skill levels of participants, etc). Although this method is well renown for its clear taxonomy of program outcomes, this method potentially provides educators with only limited support to perform a complete evaluation of the professional development program due

to its linear evaluation nature (Bates 2004).

The CIPP (**C**ontext **I**nput **P**rocess **P**roduct) evaluation model (Datta 2007) places emphasis on program improvement more than focusing on the program effectiveness. The four stage evaluation process aligns well with all phases of education program: planning, implementation, and a final retrospective assessment. While the input, process and context analysis levels focus on possible improvements to the program, only the last element of this model places emphasis on program outcome evaluation. The aim of the CIPP model attaching importance to process evaluation is to look into all the strategies and components of evaluation. This model seeks to find out if evaluation design is functioning properly, what aspects of the program are the problematic ones and how they can be improved, and what are the most efficient data collection approaches with respect to the mentioned questions (Hakan & Seval 2011).

Guskey's model (2000) includes five critical levels of evaluation including participants' reactions reaction, participants' learning, organisational support and learning, use of new knowledge and skills, and student learning outcomes. The reaction level analysis evaluates the participants' satisfaction and impressions about professional learning program as a whole with respect to different aspects of it including timing, presenters, content, etc. The learning evaluation focuses on measuring gained knowledge and skills improvements of the participant. The third level, which is the key difference between Guskey and Kirkpatrick evaluation models, focused on evaluating how the organisation supports and reinforces the newly acquired skills and knowledge. Level four of Guskey's model focuses on examining participant's use of newly gained knowledge and improved skills in the classroom environment. Finally, the student learning outcome evaluation phase focuses on understanding the extension to which the students learning outcome has been significantly improved as a result of the teacher professional development program. Guskey placed Kirkpatrick's four stage model as the basis of his work and altered it in a way that it perfectly matches to educational implementations. As a result, his evaluation model has attracted a considerable amount of attention in academia.

The interconnected evaluation model (Clarke & Hollingsworth 2002) focuses on different paths that professional development can take. This model is based on the assumption that *change* happens mainly through two processes: *reflection* and *enactment*. Reflection refers to active consideration leading to inferences that causes change in beliefs and practice. Enactment means attempting a new practice or putting in practice what is learnt and the changed belief. Enactment and reflection, according to interconnected model, occur in four domains: the personal domain (teacher knowledge, beliefs and attitudes), the domain of practice (new teaching style, modified pedagogy), the domain of consequence (outcomes such student learning outcome and engagement increase), and the external domain (sources of information, stimulus or support, workshop, webinar, etc). Evaluation of a professional development program based on the

interconnected model involves measuring changes in aspects of these domains. For example, if, as a result of attending a training course, the teacher tries a new teaching strategy in the classroom, then there has been a change in the practice domain. This model is named the interconnected model since its core concept is based on the fact that change in one domain can be observed in all other domains through the two processes of reflection and enactment.

One of the concerns with Kirkpatrick's evaluation model is that it does not take other contributing factors which can influence the learning outcomes of the students. For example, if a teacher attempts a professional development program but there is no significant change in the learning outcome after completion of the professional development, a Kirkpatrick evaluation of the professional development program would rate the program as unsuccessful. However it could have been the case that the organisation in which the teacher practices does not provide the right support and has a poor support for the requirements of the new practice for the teacher. Holton's HRD (**H**uman **R**esource **D**evelopment) evaluation model (Holton III 2005) is specifically designed to detect such casual influences of professional development outcomes. Evaluation of a professional program using HDR model is performed in three different levels including learning (knowledge, skills and attitudes), individual performance (new teaching strategies, improved teaching style, better understanding of pedagogy), and organisational performance (student outcomes). Holton proposes a set of secondary influences which can explicitly or implicitly impact the professional development outcomes on learning, and/or individual performance levels.

Desimone (2009) introduced a critical feature based measurement for analysing the impact of teacher professional development interventions. This conceptual framework for studying the effects of professional development programs on teachers and students includes four levels characterised by critical features. These levels include core features of professional development, changes in knowledge, skills, attitude, and/or beliefs, change in instruction, and changes in student learning. The first layer of evaluation focuses on some of the characteristics of teacher professional development programs which are fundamental to increasing teacher knowledge and skills that will improve teacher practice and as a result the student learning learning outcomes. These features include content focus (professional development is designed to focus on activities concerning subject matter e.g., biology, history, algebra), active learning (the professional development program provide opportunities foe the teachers to engage in active learning), coherence (the extent to which teacher learning is consistent with teachers' knowledge and beliefs), duration (sufficient number of hours spent during the program and the span of time that the activity spreads on), and collective participation (participation of teachers from the same school, grade, or department). The philosophy behind this model is based on the finding of previously published works (Desimone et al. 2002, Garet et al. 2001) which suggest that analysing a teacher learning opportunity as a measurable phenomena should be based on measuring a set of characteristics that make the profes-

sional learning program effective for increasing teacher learning and changing practice, and ultimately for improving student learning rather than on the type of activity. This model provides the opportunity to investigate the results of the professional development program not only from a teacher change theory perspective, but also on a theory of instruction level.

Based on the literature above, there has been a considerable amount of propositional work in the research area of teacher professional development evaluation. However, the proposed frameworks are not based on any evidence that one model is more effective than any other. As well, many workshop evaluations do not utilise established models and metrics to assess their efficacy, and often do not draw upon theory to situate their work (see review below). Moreover, there has not been a systematic review of teacher professional development workshops, despite the fact that workshops for teachers form a fundamental part of the teacher education landscape. Understanding how teacher professional development workshop have been evaluated and researched in the past enables educators and researchers to derive recommendations for how teacher professional development workshop evaluations can be improved in future. Also, there are no professional learning evaluation models that relate specifically to technology training. However, there has been some extensive work on evaluating technology use in education (Lai & Bower 2019). That study conducted a systematic review of how technology use in education has been evaluated, based on 365 papers published in Computers and Education between 2015 and 2017. They report that the evaluation of learning technology use tended to focus on eight themes: learning outcomes, affective elements, behaviours, design, technology elements, pedagogy, presence, and institutional environment. These eight evaluation dimensions have been used as the basis of the current study's technology evaluation framework, because it provides a comprehensive and evidence-based foundation for conceptualising evaluation areas.

4 Methods

This section explores how the set of 41 publication were selected from the literature. The systematic review protocol used in this publication aligns with the Systematic Reviews and Meta-Analyses (PRISMA) methodology (Moher et al. 2009). The five stages in this systematic review were: 1) literature search, 2) article selection, 3) data extraction, 4) data analysis, and 5) data synthesis.

4.1 Publication Selection Process

In order to find publication which suit the purpose of this study, the search was performed in June, 2020, and the focus was put on three main publication search databases including Web of Science, Scopus, and Education Resources Information Centre (ERIC) (See Figure 1). The most widely know search strategy is

the use of electronic search engines (Wolery & Lane 2014). To complement the search, five journal venues which are highly valued in education research including *Teaching and Teacher Education journal*, *journal of Teacher Education*, *Professional Development in Education*, *Technology, Pedagogy and Education*, and *Australian journal of Teacher Education* were also searched. The search included the range from January 2000 to June of 2020 so that the nature of the workshops and the sorts of technologies used were at least somewhat relevant to contemporary teaching and evaluation processes. The Boolean ("OR" and "AND") and wildcard (*) search terms used to find publications in the above-mentioned databases were as follows: "teacher" OR "educator" AND "workshop" OR "webinar" AND evaluat* AND technolog*. The full search strings are provided in Appendix A. Searches were defined to return any publication with the appearance of the mentioned words in their title, abstract, or descriptors. This search process resulted in a total number of 445 documents. Out of these 445 documents, 52 documents were detected by more than one search engine hence the duplicated documents were removed from the document pool. Upon closer inspection of the remaining 393 documents, it was noticed that six of these document were in a non-English language, another six documents were detected under slightly different names or from other publishing venues. As well, 36 documents were calls for conference participation, books, workshop calls, or book chapters. These 48 documents were also removed from the document pool because they had not undergone a peer review process or did not involve empirical work.

An Excel spreadsheet was used to document titles and abstracts to organise the screening, exclusion and inclusion processes. The decision about whether to include the paper was based upon whether it met the following criteria: 1) the paper reported upon a workshop, 2) the paper related to technology, 3) the paper aimed to train teachers, and 4) the paper placed emphasis on evaluating training. In other words, if the paper was not focused on training teachers in a workshop, or it was not conducting any evaluation on technology training then they were being removed from further analysis.

The abstracts of the remaining 345 documents were inspected to exclude those documents which did not include teachers as subjects or did not have any emphasis on teacher training. Teachers in this study included both pre-service and in-service teachers, as well as university lecturers and tutors. However, if the paper discussed the training included both teachers and students it was not included in the analysis. We accepted webinars if the nature of the webinars included training teachers by using a technological product. Furthermore, we did not include workshops with follow-up implementation support, for instance, a mentoring program afterwards, as these were deemed to be more of a program than a workshop. This abstract screening process resulted in removal of 164 documents. In cases where it was uncertain whether inclusion or exclusion was appropriate, the main body of the documents were inspected in the next phase.

Next the focus was narrowed down to those documents which focus on teacher professional development technology workshops only. In other words, the workshops trained the teachers in using technology, but not purely using

technology as a way of demonstrations. This was necessary because some of the documents were reporting purely on teacher's feedback on a specific approach or a technological product. Another exclusion criterion was disregarding the professional development that involved a broader program with multiple components (e.g. information sessions, mentoring program, etc.) rather evaluation of workshops only.

In order to assess the veracity of the paper selection process, a second member of the research team independently undertook the selection process using the protocolos above. Training involved screening approximately 5 papers together, with the two researchers comparing results and had discussions to make sure that there was a consensus on the nature of teacher training (i.e. teachers trained by using a technological product) in the papers. After the full-text assessment against the criteria, inter-rater agreement (IRA) of the two authors was conducted by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100 (Allen et al. 2020) and the IRA of the screening phase was 97%. The disagreements were discussed and resolved which a total number of 41 papers was included in the analysis.

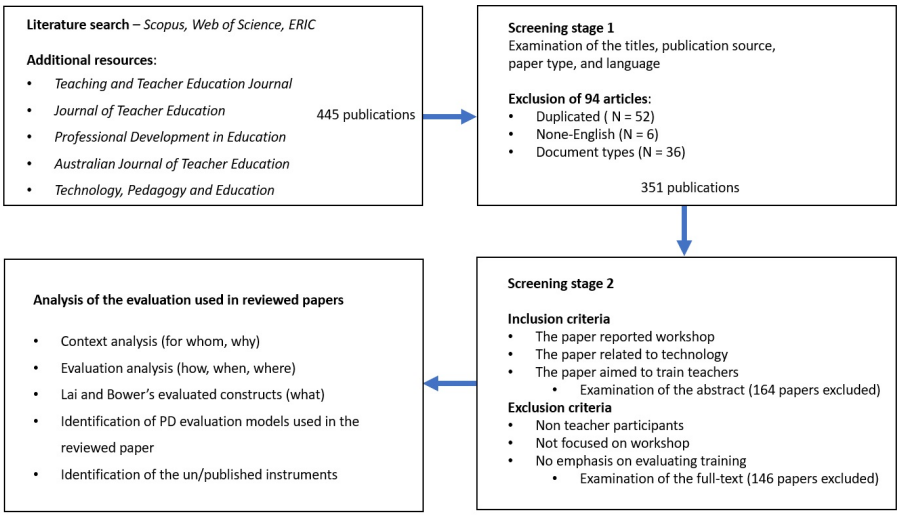

Figure 1: A diagrammatic representation of the literature selection process.

4.2 Characterising the evaluation process of the selected papers

To understand the nature of the evaluation of the workshops reviewed in this study, each paper was examined in order to find out when the evaluation has happened (before, during, straight after the workshop, or sometime after the completion of the workshop), how the evaluation has happened (surveys, inter-

views, focus group sessions, etc), the use of any of the evaluation approaches reviewed in the literature section of this paper, what had been evaluated, the research area behind each workshop, delivery mode of the workshops, participating teachers' discipline and teaching year and number of workshops held for each study. To accomplish these objectives, the soft copies of all 41 papers were obtained from the publishing venue and imported as text documents in NVivo 12 qualitative data analysis software. Each paper was read, and informative parts of the paper were coded according to the nodes which were defined to represent the aforementioned objectives. Also, an Excel spreadsheet was created to record the descriptive coding of the studies, i.e. 1) Number of workshops implemented in the study; 2) Teaching level of the participants; 3) Teaching discipline of participants; 4) Evaluation research methods used in the reviewed papers; and 5) Types of data being collected. Previous systematic literature reviews, such as that by Crompton et al. (2017) used similar approaches to record review information.

Two authors met to clarify and confirm the definitions of the descriptive coding. For instance, the number of workshops per study was confirmed to be determined by the number of repetitions of the original workshop. The teaching disciplines and research methods were divided into ten and seven categories respectively. The type of data being used in evaluation was examined, and for example, a study was classified as collecting quantitative data if more than half of the survey questions were quantitative in nature.

The main author and second member of the research team coded the 41 papers independently. In order to compare the coding results between the two authors, Correlation Coefficient (cc) analysis were performed (Taylor 1990) to examine the degree of compatibility of the number of items per coding. Relating to the number of workshops per study as well as the type of data, the categorisation and the number of papers were the same between the two authors (cc = 1). Regarding the teaching levels, the correlation coefficient was 0.87 (very high level of agreement). The categorisation of the teaching disciplines was slightly different between the two authors. For instance, the second author created a multi-disciplinary category including teaching disciplines such as English language, mathematics, social studies, health and physical education, etc. in one study (Ciampa 2016), with correlation coefficient 0.81 (high level of agreement). Finally, the categorisation of the research methods used in studies between the two authors was nearly the same (cc = 0.86) with a slightly different number of studies on a few categories.

4.3 Identification of evaluated constructs

In order to identify different aspects that are evaluated in technology workshops, the construct categorization recently published by Lai and Bower (2019) was used. The evaluated content of each paper was examined according to their scheme in order to categorise the evaluation foci in Lai and Bower framework (2019). Identification of the constructs being evaluated by each paper was based on the instruments used, the findings in the result section of the

papers, the objectives of the program, manual examination of the instruments provided in the appendixes of the papers, and the purpose behind evaluation. Lai and Bower framework (2019) was able to sufficiently characterise all of the focuses of evaluation represented in the 41 papers. Two authors independently appraised each article's evaluated contents against the eight dimensions. The results were recorded in the same spreadsheet used for the descriptive coding with eight separate columns including learning, affective elements, behaviour, design, teaching or pedagogy, technology, presence and institutional environment. Each dimension was coded dichotomously (0 = Absent in the study, 1= Present in the study). Both researchers were knowledgeable about the 2019 framework and the framework was in fact developed by Coder 2. The two authors then compared and discussed their coding, and came to an agreement on the eight themes. The correlation coefficients of the eight dimensions between the two researchers were 91% which was deemed as reliable (Taylor 1990).

4.4 Mapping evaluations to professional development evaluation models

In order to map the evaluated content of the workshops against workshop evaluation models, it was decided to consider all six professional development evaluation models reviewed in the literature section for two reasons. Firstly despite the fact that professional development evaluation methods show a high degree of convergence with respect to what they tend to evaluate, each of the six evaluation models offers a or a combination of unique feature/s to it which is either not covered by other models or not emphasised. Secondly, the fact that two of the five publications (found in the set of 41 paper examined in this study which use published professional development assessment models for evaluation purposes) use more than one evaluation model provided additional motivation to consider more than one workshop evaluation model to analyse which aspects of the evaluations performed in the set of 41 papers map to what area/s of the five professional development evaluation models. Therefore, each evaluation model was examined to create a meta-model that provides an overview of how the models evaluated professional learning, and when the evaluation happens. In general, professional development workshop evaluation models perform evaluation in four main phases including prior to workshop, during workshop, after the workshop, and longitudinal impact. Stufflebeam's input and context evaluation as well as Holton's motivational element and secondary influence analysis happen prior to the workshop. Stufflebeam's process level evaluation occurs during the workshop. Evaluations performed in Kirkpatrick's learning, reaction and behaviour levels, Guskey's learning and practice levels, Desimone's learning and behaviour levels, Stufflebeam's results phase, Clarke's personal domain and external domain, and Holton's individual performance level are performed directly following the workshop. Finally, Kirkpatrick's evaluation of results, Guskey's evaluation of results and behaviour, Desimone's evaluation of results, Clarke's evaluation in practice domain and consequence domain, and Holton's evaluation of organisational results occur sometime after the completion of the

profession development workshop (See table 5).

5 Results

5.1 Contexts being evaluated

The 41 papers included 21 journal articles and 20 peer reviewed conference papers that evaluate teacher professional development technology workshops (3 webinars, 1 blended and 37 face to face). There was a diverse range of technologies that were being studied, with the most popular technologies being 4 studies on Computer Assisted Language Learning (CALL), 4 studies on mobile devices, and 3 studies on game based learning. The other technologies observed include Assistive technology (AT), cloud computing, Computer Supported Collaborative Learning (CSCL), High Performance Computing (HPC), e-learning, digital libraries, Table PC, webinars, webinars, open source tools, and digital libraries (each paper focusing on one technology). The majority of professional development programs (68%) included only one workshop (See Table 1). The remaining studies are based on instantiation of more than one workshop with 5 repetition of the original workshop being most popular (4 studies). In two cases, the evaluation was performed on more than 10 workshops (repetition of the original workshop). Having multiple iterations of the same workshop was seen as useful because the evaluations of the early implementation usually guided the improvement plan for future workshops. As shown in Table 2 the majority of the participants of the workshops were secondary teachers (61% of studies), with four of the workshops designed specifically for university teachers, 10% targeting pre-service teachers, 7% primary school teachers, with 12% of studies targeting in-service k-12 teachers. A total number of 9 studies did not mention the target participants' teaching level. Regarding the teaching discipline of the participating teachers, 52% of the studies were specifically designed to target Science Technology Education Mathematics teachers with 9, 8, and 5 papers specifically designed for teachers active in teaching Mathematics, Science, or general STEM subjects respectively (see Table 3. Five of the studies were targeting teachers who taught languages (English, Spanish, Chinese). The rest of the workshops were specifically designed for teachers teaching either arts, music, social studies, medical or health subjects. Interestingly, only two of these papers focused on teachers teaching IT, engineering, or computer science subjects.

5.2 RQ1: Constructs examined when evaluating workshops

A total of 23 studies (56% of the papers) aimed to assess if participants have learnt something new as a result of attending the teacher technology professional development workshop (See Table 4). The majority of the studies (35 papers) attempt to understand teachers perceptions, attitudes, and intention of use of a new technology or approach. There were 16 workshop evaluation papers that

Table 1: Number of workshops implemented in studies.

No. of workshops per study	**No. of studies**	**% of the studies**
1	28	68%
2	1	3%
3	2	4%
4	2	4%
5	4	9%
7	1	3%
9	1	3%
19	1	3%
48	1	3%
98	41	100%

Table 2: The teaching level of the participants of the evaluated publications.

Teaching level of the participants	**No. of studies**	**% of the studies**
Pre-service	4	10%
Grades 1 to 6	3	7%
Grades 7 to 9	10	25%
Grades 10 to 12	15	36%
K-12	5	12%
Higher education	9	21%
Unspecified	6	14%

aimed to understand how teacher react to a tool or approach in a practical sense through providing participants access to the tool or use the approach to attempt problem solving during the workshops. A total number of 18 papers attempted to evaluate technology adoption, function, ease of use and perceived usefulness. Interestingly, only a minority of papers (12% of papers) attempted to examine changes in teacher behaviour in the workplace due to the workshop. A large number of studies (75% of papers) attempt to evaluate the design of the content presented in their workshops. Very little attention has been given to assessment of institutional implications of technology use (7% of the papers) or the impact of technology use in facilitating a stronger presence of teacher (2%), presence (2%), or communities (2%). Surprisingly, only a few studies examined participants teaching strategies (3%), teaching quality (7%), or teacher's used pedagogical practices. Lastly, self-regulation, collaboration or interaction with others, and participation of teachers with the workshop is under-evaluated in the examined studies with 4%, 4% and 9% of the entire population of studies examined in this review focusing on those aspects.

Table 3: The teaching disciplines of the participants of the evaluated publications.

Teaching discipline of the participants	**No. of studies**	**% of the studies**
Science	8	19%
Maths	9	21%
STEM	5	12%
Engineering and IT	2	4%
Languages	5	12%
Medical	1	2%
Social studies	1	2%
Health	1	2%
Arts	1	2%
Music	1	2%

Table 4: Technological constructs examined by workshop evaluations performed in the studies reviewed.

Constructs	**% of papers**	**Elements**	**No. of studies**	**%**
Affective Elements	66%	Attitudes, values or beliefs	12	29%
		Emotions like boredom joy	1	2%
		Engagement, motivation	5	12%
		Perceptions, intentions	14	34%
		Self-efficacy	11	26%
Behaviour	51%	Interaction, collaboration or cooperation	2	4%
		Self-reflection, self-evaluation	16	39%
		Self-regulation	2	4%
		Usage or participation	4	9%
Design	88%	Course quality	31	75%
		Course content	31	75%
		Course structure	15	36%
		Resources	10	24%
		Overall design	17	41%
Institutional Environment	12%	External environment factors	0	0%
		Institutional capacity	0	0%
		Institutional intervention	1	2%
		Institutional policy	1	2%
		Institutional support	3	7%
Learning	80%	Cognitive load or effort	0	0%
		Knowledge, achievement or performance	23	56%
		Learning styles or learning strategies	1	2%
		Skills development	10	24%
Presence	7%	Presence in the environment	1	2%
		Social presence,	1	2%
		co-presence or community	1	2%
Teaching or Pedagogy	32%	Feedback	3	7%
		Pedagogical practice,	1	2%
		teaching strategies	5	12%
		or teaching quality	3	7%
Technology	49%	Accessibility	2	4%
		Adoption	5	12%
		Functionality	3	7%
		Perceived usefulness	18	43%
		Perceived ease of use	8	19%

5.3 RQ2: Evaluation frameworks used in studies that evaluate educational workshops

Two papers base their evaluation method on Guskey's evaluation model (Guskey 2002*a*), one paper that deploys Kirkpatrick based evaluation (Kirkpatrick & Kirkpatrick 2006), and two papers that use a mix of the two mentioned evaluation methods. To summarise, only 5 papers out of 41 papers were found which based their evaluation framework on a theorised professional development evaluation model. This aligns with previous findings by Patton that formal professional learning evaluation frameworks are seldom used researchers (Patton 2008). However, one needs to keep in mind that not highlighting an evaluation model in a published work does not mean that the evaluation performed is incomplete.

None of the 41 investigated papers took full advantage of any of the evaluation models. Specifically, the few studies that did mention use of the Kirkpatrick or Guskey model did not fully utilise these models and only took some of the evaluation framework into account. According to Table 5, Stufflebeam's CIPP seems to be the evaluation model with most number of areas found to be related to workshop evaluation programs. This is due to three main reasons. Firstly, the majority of the papers investigated in this study have implemented workshops as professional development program without any attempt to investigate the results of the workshop at teacher's workplace level which aligns well with the nature of CIPP which focuses on the program itself rather than it future implications. Stufflebeam's approach is the only evaluation model which considers collection and analysis of participants data prior to the workshop for program improvement purposes. This aligns well with the sort of changes that organisers of an educational workshop aim to make to the program based on teacher needs, characteristics, level of competence and expectations. Lastly, Stufflebeam's model satisfies organisational requirements of evaluation very well since it mainly focuses on program evaluation without centrally focusing on the participants. Holton's evaluation model maps the least to the evaluation approaches of the workshops since it mainly focuses on secondary influences that can impact upon learning, practice, and organisational results.

As can be interpreted from Table 5, workshops investigated in this paper attempt to collect information from the participants prior to the intervention. According to our analysis of the papers, the main reason for collection of data prior to the workshop was to understand the number of participants, their requirements, and to collect background information. None of the studies, however, report changes to the development of the program due to the analysis of such data. One of the studies reported that collection of the data prior to the program had mainly been performed for the purpose of selection of the workshop presenters and preparation of the intervention with accordance to participants level of competence. Despite the fact that most workshops attempt to collect data prior to the intervention, the main goal of the data collection has been to assist with comparing the state of the participants after the workshop and not

Table 5: Professional development evaluation frameworks used in evaluation of teacher professional development workshops.

Author	Pre		During		Post		Longitudinal Impact	
Holton	secondary influences motivation elements environmental elements learning design ability				motivational elements performance influences content validity transfer design transfer capacity opportunity to use	✓	external events expected utility organisational work	 ✓ ✓
Desimone					learning→behaviour	✓	learning outcome	✓
Kirkpatrick					workshop learning attitudes/perceptions	✓ ✓ ✓	learning outcome	✓
Guskey					learning→behaviour →student benefits	✓	→ attitude/perception	
Clarke					workshop/satisfaction	✓	attitude/perception learning outcomes curricular development new teaching strategies	✓ ✓ ✓ ✓
Stufflebeam	background info needed interventions teacher competency teacher needs planning teacher characteristics	 ✓ ✓ ✓ ✓	attitudes participation satisfaction	✓ ✓	learning attitudes/perceptions program	✓ ✓ ✓		

for the sake of program development purposes. We also did not notice an attempt to analyse participants personality traits and other secondary influences which can impact learning and motivation to change.

5.4 RQ3: prevailing instruments that are used to evaluate educational workshops

Teacher professional development workshops use a variety of techniques to collect data from the participants prior to the workshop, during the work or after the workshop. Based on the findings, the dominant approach to collect participant data is questionnaires or surveys. A total of 21 studies (51% of the papers) used pre-test post-test approach in order to detect traces of teacher change due to attending the professional development program. Two of the studies only analysed the data collected prior to the workshop and used other types of data collection techniques during the workshop. Surprisingly, a total number of 18 papers only analysed the data collected post-workshop without performing pre-workshop data collection. Ten of the workshops attempted to collect data during the workshop using different data collection techniques including log files, verbal feedback, video recordings, interviews, direct observations and written reflections. Only six studies attempted to communicate with the teacher an extended time after completion of the workshop to examine changes in teacher behaviour or workplace outcomes due to impact of the professional development program.

Based on our findings presented in Table 8, majority of the studies (75%)attempted to evaluate participant's satisfaction (affective reaction) to the workshop. This was typically achieved using custom made opinion forms and satisfaction surveys (marked 'NA' in Table 8). In fact, amongst all of the instruments used to evaluate workshops, there was only 7 cases amongst the sample of 41 studies where previously published instruments were used. Three of these pa-

pers related to evaluation of TPACK ((Graham et al. 2009), (Hu & Fyfe 2010) and (Schmidt et al. 2009)), with the other four evaluating self-efficacy (Canbazoğlu Bilici 2013) , attitude (Kennedy-Clark 2011), tool usefulness (Carroll et al. 2009), and opinion of the teacher on argumentation test (Sampson & Clark 2006). Three of these tests (TPACK evaluation surveys) were used in the same study, with the other four used in four separate papers. To summarise, only 12% of the investigated studies used a published instrument to perform evaluation.

Table 6: Research methods used in the reviewed papers to evaluate the use of technology in education.

Research methods used in studies	**No. of studies**	**% of the studies**
Quasi-experiments	4	10%
Experiments	14	34%
Surveys	34	82 %
Case studies	5	12%
Interviews or focus groups	12	29%
Design-based research	3	7%
Observations	11	26%

Table 7: Types of data being collected.

Types of data	**No. of studies**	**% of the studies**
Quantitative (including Likert scale)	20	49 %
Mixed method (both quantitative and qualitative)	18	44%
Qualitative	3	7%
Total	41	100%

6 Discussion

Five out of 41 papers included theoretical framing, with Kirkpatrick evaluation model and Guskey evaluation framework being most popular. Only 9 papers conceptualised workshops in terms of teacher change, which is arguably the most established and robust approach to analysing technology professional development workshops. Based on the literature review provided in this study, there is a wide range of professional development evaluation frameworks that can be used to assess teacher professional development workshops. These approaches provide thorough methodology to perform evaluation in different stages based on different evaluation objectives. Each framework however has a unique signature with respect to its application which can be used for a particular purpose. Kirkpatrick's model aims to evaluate workshop outcomes with respect to participant satisfaction, learning, and results of professional development measured as student outcomes. Guskey puts the teacher at the centre of professional development and assesses a professional development workshop successful if there is

Table 8: Instruments used in studies.

Instrument / Element	Construct	Reference	No. of papers
Attitude towards technology / Attitude		NA	5
Technology use / Beliefs		NA	5
Individual learning forms / Value		NA	4
Computer science (technology) / Knowledge		NA	2
Team learning form / Value		NA	1
Argumentation test / Attitude		(Sampson & Clark 2006)	1
Technological pedagogical content / Self-Efficacy	Affective Elements	(Canbazoğlu Bilici 2013)	1
Knowledge self-efficacy belief scale / Self-Efficacy		NA	1
Perspective form / Attitude		(Kennedy-Clark 2011)	1
Technology use / Self-efficacy		NA	1
Creative support index / Perceptions		(Carroll et al. 2009)	1
Classroom use / STEM knowledge		NA	1
Technology use / Usage	Behaviour	NA	4
Teacher's confidence in TPACK / Self-evaluation		(Graham et al. 2009)	1
General workshop evaluation forms / Design	Design	NA	31
Perceptions / School policy	Institutional Environment	NA	2
Project based learning / Learning Styles	Learning	NA	2
Individual knowledge form / Discipline Knowledge		NA	2
Technology (CS;TPACK) / Perceived Usefulness		NA	6
TPACK evaluation / Perceived Usefulness	Technology	(Hu & Fyfe 2010)	1
TPACK evaluation / Perceived Usefulness	Technology	(Schmidt et al. 2009)	2
Software evaluation / Functionality		NA	2

a change in teacher's values as a result of improved students' learning outcomes. Placing the workshop in the centre, CIPP evaluates the workshop and its different aspect from a project success point of view. Holton puts the focus on influences which implicitly or explicitly impact teacher's learning, performance, and institutional outcomes. Emphasising improvement of students' learning outcomes, Desimone evaluates the effectiveness of a teacher professional development program based on changes in teacher behaviour as a result of the professional development program. Clarke's interconnected evaluation model looks at evaluation from a non-linear perspective where teacher's satisfaction, changes in knowledge and skills, new teaching practices and improved students' learning outcome altogether are considered to evaluate the program.

Table 5 provides an overview of all professional learning evaluation models that educators and researchers can use in synthesis to guide their evaluation processes. Each of the phases of the selected evaluation framework need to be examined by the workshop organisers in order to make the evaluation of the professional development program thorough and complete. To achieve that, data needs to be collected from the participants prior, during and after the workshop complemented by collection of data in after execution of workshop to assess changes in teacher behaviour and longitudinal impact. Based on the finding of this study, only 58% of the papers investigated in this study perform data collection on a pre-post workshop level, 4 studies only collected data prior to the workshop, and 24 studies that collected data post-workshop did not collect any data from the participants prior to the workshop. The analysis of the impact of the professional development as a result of attending the workshop requires assessing changes observed which can only be accomplished by comparing the pre- and post-workshop data. Furthermore, it was observed that only 9 (21%) of the papers attempted to analyse the impact of the longitudinal impact of the

workshop.

With respect to what is evaluated, based on the findings of this review most studies only involve a small subset of applicable dimensions in their evaluation. 50% of workshops focus on assessing changes in teacher values, 87% evaluate workshop design and outcomes, and 9 papers aim to assess the usefulness of a tool or an approach. It was observed that those elements relating to different pedagogical implications are not fully investigated. Only 6 papers investigated institutional elements, presence, changes in teacher behaviour, changes in teaching strategies and improvements in students learning outcome as a result of the teacher professional development workshop. It is recommended that a more comprehensive approach to evaluation of teacher professional development workshop is used that involves many, if not all, eight evaluation dimensions reviewed in Table 4 .

As mentioned in the results section of this paper, the majority of the workshops (83%) did not use established instruments in order to collect data from the participants. There are a number of instruments that could be used to evaluate different aspects of professional learning workshops, for instance, Mental Effort Scale (Mulder 1986) derived from Cognitive Load Theory to evaluated learning or Study Processes Questionnaire (SPQ) (Biggs et al. 2001) can be used. Technological Pedagogical Content Knowledge (TPACK) framework (Koehler & Mishra 2009) can be used to evaluate learning in technological contexts. Evaluation of affective elements can be done using Self-efficacy Scale (Sherer et al. 1982), Achievement Emotions Questionnaire (Pekrun et al. 2011), Positive and Negative Affect Schedule (Watson & Clark 1999) or Motivated Strategies for Learning Questionnaire (Pintrich et al. 1991). Online Self-Regulated Learning Questionnaire (OSLQ) (Winne & Perry 2000) can be used for evaluation of behaviour. With respect to the technology itself, Technology Acceptance Model (TAM) (Szajna 1996), Diffusion of Innovation Theory (Kaminski 2011), Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al. 2016), Expectation–Confirmation Model (Bhattacherjee 2001) and System Usability Scale (SUS) (Brooke 1986) are among commonly used instruments that are used to evaluate different aspects of the technology (Lai & Bower 2019). Using a full survey can be impractical as completion of one or more of these surveys could be time consuming or burdensome for participants, and in some cases not all questions provided in such surveys are needed. The authors of this paper recommend that workshop organisers at least review established instruments to ascertain the applicability of various items for their study.

Many of the studies (93%) did not provide the full content of the instruments used, focus group questions, or evaluation material used to assess different aspects of the workshop. Having no access to the used instruments makes it unclear to grasp what aspects of the teacher professional development workshop are evaluated, which makes it hard to replicate the evaluation approach for subsequent comparison and contrast. It is also important to note that 90%

of studies attempt to evaluate multiple constructs using the same instrument and report only on a small portion of the results of the survey. Most studies limit the reporting of statistical findings from the instruments used, with only 11 studies attempting to empirically reflect on the findings in an educational setting. This is to some extend understandable since nearly half of the papers investigated in the current study aim to report on the overall program success rather than focusing on teacher change and the implications of this change on the teacher or the student. However, if teacher professional learning workshops on technology use are going to involve rigorous research as a sub-field of teacher education, then a more robust methods (e.g. pre-post, qualitative analysis e.g. from interviews) needs to be adopted.

7 Conclusion

In this paper a systematic literature review of the evaluation models and approaches used for teacher professional development workshops for the use of technology was performed. The majority of the evaluations performed in the 41 reviewed papers do not utilise evaluation frameworks which are designed to assess the professional development workshop programs. Most of the evaluation relates to assessing workshop and the effect of workshops on the beliefs, attitudes and behaviour of participants, rather than other elements such as impact on pedagogy, design practices, presence and community, or institutional factors. Almost none of the workshops performed follow up studies to examine the impact of attending such workshops upon teacher workplace practices, let alone student learning. Most instruments used in these workshops are designed by the workshop facilitators and are not based on standard established instruments that measure different aspects of workshops.

While this systematic review successfully addresses its pre-identified research questions, it also comes with some limitations. One of the important limitation of this study is that not all evaluations of technology integrating workshops are peer-reviewed and published or presented, so the findings only relate to workshop evaluations published in research articles. We contend that workshops presented in peer-reviewed research would tend to be more rigorously evaluated than workshops evaluations in non-peer-reviewed publications, conducted by commercial entities, or not published at all, however, further research would be needed to establish this.Another important limitation is that non-workshop professional development programs may tend to be evaluated better, for instance those occurring over extended durations. For this reason it is important to note that our findings only apply to workshops.

Workshop evaluations often need to be expedient, for pragmatic reasons. However, if workshops are to be evaluated for research purposes, then we argue that those evaluations need to be a) grounded in the theory of workshop evaluation and teacher change, b) consider a comprehensive range of educational dimensions to evaluate, c) collect a variety of data before, during and

after workshops, and d) refer to established instruments rather than relying entirely on self-created and theoretically agnostic questionnaires. Only then can researchers form a more complete understanding of the impact of various workshop approaches and progress the science of professional learning.

References

Allen, G. E., Common, E. A., Germer, K. A., Lane, K. L., Buckman, M. M., Oakes, W. P. & Menzies, H. M. (2020), 'A systematic review of the evidence base for active supervision in pre-k–12 settings', *Behavioral Disorders* **45**(3), 167–182.

Bates, R. (2004), 'A critical analysis of evaluation practice: the kirkpatrick model and the principle of beneficence', *Evaluation and program planning* **27**(3), 341–347.

Bhattacherjee, A. (2001), 'Understanding information systems continuance: an expectation-confirmation model', *MIS quarterly* pp. 351–370.

Biggs, J., Kember, D. & Leung, D. Y. (2001), 'The revised two-factor study process questionnaire: R-spq-2f', *British journal of educational psychology* **71**(1), 133–149.

Brooke, J. (1986), 'System usability scale (sus): a quick-and-dirty method of system evaluation user information', *Reading, UK: Digital Equipment Co Ltd* **43**.

Bruce, C. (2003), Information literacy as a catalyst for educational change: A background paper, *in* 'International Information Literacy Conferences and Meetings', NCLIS. gov, pp. 1–17.

Canbazoğlu Bilici, S. (2013), 'Fen ve teknoloji öğretmenlerine teknolojik pedagojik alan bilgisi kazandırma amaçlı eğitim uygulamaları ii'.

Carroll, E. A., Latulipe, C., Fung, R. & Terry, M. (2009), Creativity factor evaluation: towards a standardized survey metric for creativity support, *in* 'Proceedings of the seventh ACM conference on Creativity and cognition', pp. 127–136.

Ciampa, K. (2016), 'Implementing a digital reading and writing workshop model for content literacy instruction in an urban elementary (k–8) school', *The Reading Teacher* **70**(3), 295–306.

Clarke, D. & Hollingsworth, H. (2002), 'Elaborating a model of teacher professional growth', *Teaching and teacher education* **18**(8), 947–967.

Cook, D. A. (2010), 'Twelve tips for evaluating educational programs', *Medical teacher* **32**(4), 296–301.

Crompton, H., Burke, D. & Gregory, K. H. (2017), 'The use of mobile learning in pk-12 education: A systematic review', *Computers & Education* **110**, 51–63.

Datta, L.-e. (2007), 'Evaluation theory, models, and applications, by daniel l. stufflebeam and anthony j. shinkfield. san francisco: Jossey-bass, 2007. 768 pp.', *American Journal of Evaluation* **28**(4), 573–576.

Desimone, L. M. (2009), 'Improving impact studies of teachers' professional development: Toward better conceptualizations and measures', *Educational researcher* **38**(3), 181–199.

Desimone, L., Porter, A. C., Birman, B. F., Garet, M. S. & Yoon, K. S. (2002), 'How do district management and implementation strategies relate to the quality of the professional development that districts provide to teachers?.', *Teachers College Record* **104**(7), 1265–1312.

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E. & Sendurur, P. (2012), 'Teacher beliefs and technology integration practices: A critical relationship', *Computers & education* **59**(2), 423–435.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F. & Yoon, K. S. (2001), 'What makes professional development effective? results from a national sample of teachers', *American educational research journal* **38**(4), 915–945.

Graham, R., Burgoyne, N., Cantrell, P., Smith, L., St Clair, L. & Harris, R. (2009), 'Measuring the tpack confidence of inservice science teachers', *TechTrends* **53**(5), 70–79.

Guskey, T. R. (2000), *Evaluating professional development*, Corwin press.

Guskey, T. R. (2002*a*), 'Does it make a difference? evaluating professional development', *Educational leadership* **59**(6), 45.

Guskey, T. R. (2002*b*), 'Professional development and teacher change', *Teachers and teaching* **8**(3), 381–391.

Hakan, K. & Seval, F. (2011), 'Cipp evaluation model scale: development, reliability and validity', *Procedia-Social and Behavioral Sciences* **15**, 592–599.

Holton III, E. F. (1996), 'The flawed four-level evaluation model', *Human resource development quarterly* **7**(1), 5–21.

Holton III, E. F. (2005), 'Holton's evaluation model: New evidence and construct elaborations', *Advances in developing human resources* **7**(1), 37–54.

Hu, C. & Fyfe, V. (2010), 'Impact of a new curriculum on pre-service teachers' technical, pedagogical and content knowledge (tpack)', *Proceedings Ascilite Sydney* pp. 184–189.

Kaminski, J. (2011), 'Diffusion of innovation theory', *Canadian Journal of Nursing Informatics* **6**(2), 1–6.

Kennedy-Clark, S. (2011), 'Pre-service teachers' perspectives on using scenario-based virtual worlds in science education', *Computers & Education* **57**(4), 2224–2235.

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

Koehler, M. & Mishra, P. (2009), 'What is technological pedagogical content knowledge (tpack)?', *Contemporary issues in technology and teacher education* **9**(1), 60–70.

Kostiainen, E., Ukskoski, T., Ruohotie-Lyhty, M., Kauppinen, M., Kainulainen, J. & Mäkinen, T. (2018), 'Meaningful learning in teacher education', *Teaching and Teacher Education* **71**(April).

Lai, J. W. & Bower, M. (2019), 'How is the use of technology in education evaluated? a systematic review', *Computers & Education* **133**, 27–42.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P. et al. (2009), 'Preferred reporting items for systematic reviews and meta-analyses: the prisma statement', *PLoS med* **6**(7), e1000097.

Mulder, G. (1986), The concept and measurement of mental effort, *in* 'Energetics and human information processing', Springer, pp. 175–198.

Patton, M. Q. (2008), *Utilization-focused evaluation*, Sage publications.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P. & Perry, R. P. (2011), 'Measuring emotions in students' learning and performance: The achievement emotions questionnaire (aeq)', *Contemporary educational psychology* **36**(1), 36–48.

Pintrich, P. R. et al. (1991), 'A manual for the use of the motivated strategies for learning questionnaire (mslq).'.

Poplin, C. J. (2003), 'Models of professional development.(seeds of innovation)', *THE Journal (Technological Horizons in Education)* **30**(11), 38.

Sampson, V. & Clark, D. (2006), The development and validation of the nature of science as argument questionnaire (nsaaq), *in* 'Annual Conference of the National Association for Research in Science Teaching, San Francisco, CA'.

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J. & Shin, T. S. (2009), 'Technological pedagogical content knowledge (tpack) the development and validation of an assessment instrument for preservice teachers', *Journal of research on Technology in Education* **42**(2), 123–149.

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B. & Rogers, R. W. (1982), 'The self-efficacy scale: Construction and validation', *Psychological reports* **51**(2), 663–671.

Szajna, B. (1996), 'Empirical evaluation of the revised technology acceptance model', *Management science* **42**(1), 85–92.

Taylor, R. (1990), 'Interpretation of the correlation coefficient: a basic review', *Journal of diagnostic medical sonography* **6**(1), 35–39.

Venkatesh, V., Thong, J. Y. & Xu, X. (2016), 'Unified theory of acceptance and use of technology: A synthesis and the road ahead', *Journal of the association for Information Systems* **17**(5), 328–376.

Watson, D. & Clark, L. A. (1999), 'The panas-x: Manual for the positive and negative affect schedule-expanded form'.

Winne, P. H. & Perry, N. E. (2000), Measuring self-regulated learning, *in* 'Handbook of self-regulation', Elsevier, pp. 531–566.

Wolery, M. & Lane, K. L. (2014), Writing tasks: Literature reviews, research proposals, and final reports, *in* 'Single Case Research Methodology', Routledge, pp. 50–84.