



How can we measure the impact of a
PD activity?
Ready-to-use guidelines



Information about the report/IO

IO N°2

Publication date: [30/06/2019]

How can we measure the impact of a PD activity? Ready-to-use guidelines

Project Information

Grant no. 2016-1-DE03-KA201-023103

Project title: European Network of STEM PD Centres

Project acronym: STEM PD Net

Start date of project: 01/09/2016

Duration: 36 months

Programme: Erasmus+, Key Action 2 (KA2) – Strategic Partnerships

Contact Information

Coordinating Institution: University of Education Freiburg, International Centre for STEM Education (ICSE)

Coordinator: Prof. Dr. Katja Maaß

Project Manager: Elena Schäfer, Laura Wanckel

Lead partner for this report/IO: PH FR

Website: <http://stem-pd-net.eu/>

© STEM PD Net project (grant no. 2016-1-DE03-KA201-023103) 2016-2019, lead contributions by International Centre for STEM Education (ICSE) at the University of Education, Freiburg. CC-NC-SA 4.0 license granted.



These guidelines are based on the work within the project European Network of STEM PD Centres (STEM PD Net). Coordination: Prof. Dr. Katja Maaß, International Centre for STEM Education (ICSE) at the University of Education, Freiburg. Partners: SOU Lyuben Karavelov, Koprivshitsa, Bulgaria; Prezidento Valdo Adamkaus Gimnazija, Lithuania; Ministry of National Education, Kizilay-Ankara, Turkey; Texas Instruments Education Technology GmbH, Freising, Germany; Institute of Mathematics and Informatics at the Bulgarian Academy of Science, Sofia, Bulgaria; Ugdymo Pletotes Centras, Vilnius, Lithuania; Universität Innsbruck, Innsbruck, Austria; Linköping University, Linköping, Sweden; Ministerio de Educación, Cultura y Deporte de España, Madrid, Spain; Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria; University of Gothenburg, Gothenburg, Sweden; Hacettepe University, Ankara, Turkey; Universität Duisburg-Essen, Essen, Germany.

The authors were Prof. Dr. Katja Maaß, Katharina Flößer, Anika Weihberger, Ian Galloway, Stefan Zehetmeier, Peter Nyström, Oliver Wagener, Toni Chehlarova, Georgi Gachev.

The project European Network of STEM PD Centres (STEM PD Net) has received co-funding from the Erasmus+ programme of the European Union.

The creation of these resources has been co-funded by the Erasmus+ programme of the European Union under grant no. 2016-1-DE03-KA201-023103. Neither the European Union/European Commission nor the project's national funding agency PAD are responsible for the content or liable for any losses or damage resulting of the use of these resources.

Contents	
Executive Summary	2
1 Developing one’s own concept of evaluation	2
1.1 Objectives and measurable indicators	2
1.2 Ethical issues.....	3
1.3 A possible model for evaluating PD programmes	4
1.4 Overview of evaluation methods	5
1.5 Designing the evaluation concept	7
2 Evaluation methods.....	8
2.1 Standard methods	8
2.1.1 <i>Satisfaction questionnaire</i>	8
2.1.2 <i>Teachers’ questionnaire</i>	10
2.1.3 <i>Semi-structured interviews</i>	13
2.1.4 <i>Learning Diaries</i>	15
2.1.5 <i>Classroom observations</i>	18
2.1.6 <i>Students’ questionnaires or interviews</i>	21
2.2 Innovative methods.....	24
2.2.1 <i>Bringing the best three tasks</i>	24
2.2.2 <i>Presentations of lessons in the seminar</i>	26
2.2.3 <i>Response Systems/ Digital response systems (as real time evaluation)</i>	28
2.2.4 <i>SWOT Analysis</i>	30
3 Outlook.....	32
References.....	33

Executive Summary

Most persons designing, running or organizing professional development (PD) activities, be it PD coordinators, administrators at all levels, instructors and other interested practitioners, are interested in the impact of their PD activities on teachers and in turn on their students.

The intention of this guide is to provide PD practitioners with feasible methods to evaluate their PD activities. It will outline several methods appropriate to different approaches to PD as well as their pros and cons. To support the PD practitioner the guide first gives some general information on important steps to be carried out, when aiming at an evaluation of a PD course. First, we need to define clearly defined objectives of the PD course, second we need to operationalize these aims to produce measurable indicators and then based on this we can develop the evaluation design. To facilitate this step, we will introduce a possible model for evaluating PD programmes referring to different levels of evaluation.

Once we have drafted an evaluation design we need to consider ethical issues. They are pivotal to any kind of evaluation. Not considering ethical issues may lead to serious difficulties, if teachers, students or parents complain about the data collection. Once this is clarified, decisions on the concrete evaluation methods can be made. To support this step, we will have a closer look at the different evaluation methods. These descriptions of the evaluation methods contain a short description of the method, examples, valuable pragmatic tips to be taken care of when actually using the methods and for avoiding mistakes as well as a list of pros and cons.

1 Developing one's own concept of evaluation

1.1 Objectives and measurable indicators

The impact of professional development (PD) initiatives must be measured to the extent to which PD fulfills the intended objectives. Therefore first of all, defining clear objectives for a PD initiative is important. If the overall goal of PD is to encourage teachers to do more inquiry in class, etc., what are the specific goals that will help them to do this?

Secondly, specific indicators for measurable changes are needed. What are, for examples, participants of a PD activity introducing inquiry-based learning as a teaching concept supposed to do after the course? In order to avoid producing masses of data without using them, it is vital to start the process by writing down the aims of the PD course and these aims need to be operationalized in order to be measurable. It is important to reflect critically on the aims. If these aims are too high or general they may not be measurable. Lesser aims might be easier to measure. Also, it might be wise to phrase the aims in a kind of structured, increasing chain, as illustrated in the following chain of aims for a PDPD course on cultural differences in STEM subjects. (1) The teachers participate regularly in the course; (2) they do their homework; (3) they develop an awareness for cultural differences; (4) they consider it important to deal with cultural diversity in class; (5) they use different methods to deal with cultural diversity in class.

Some aims may be harder to evaluate than others. Also, the scale of the PD programme, in time, number of participants and cost, affects what should and could be evaluated. However, despite the fact that contextual conditions frame the evaluation, ideally the aims and learning objectives of a PD programme need to be covered by the evaluation.

Once the objectives are defined and the indicators are set up, the evaluation design needs to be planned. It is beneficial to do this long in advance of a PD course as there are many different phases for measurement (before, during and after a PD course) and many different approaches for measurement (see part 2). Which design for evaluation is chosen depends on the objectives and indicators of

the PD course, the insight we want to gain (e. g. effects on teachers or their students), the duration of the course, the number of participants in the PD activity and time available for the evaluation. Apart from this it is also important to consider ethical issues, which also might influence the choice of the evaluation method.

1.2 Ethical issues

Before starting with any evaluation it is vital to take ethical considerations into account. This includes for example issues of anonymity, informed consent, data protection, regional and national ethical and legal guidelines. It is imperative to adhere to national legislation and practice in designing and executing methods for evaluation.

Everyone (e.g. teachers, students, parents) involved in the evaluation of a PD programme in any way needs to be fully informed about the purpose of the evaluation, its benefits and risks, its possible impact, and of their right to withdraw from the evaluation. Also all persons from whom data are collected, should be informed about the data storage and their rights to access it. This can be done for example through an information sheet made available to all persons concerned.

Furthermore, each person from whom data are collected will have to give unambiguous and voluntary consent before the data are used. For this, an information sheet and consent form can be used. All teachers, students (where they are mature enough) and parents (where students are not mature enough) need to be asked to read and sign the consent form prior to the data collection.

In classrooms, where children are involved, you should comply with Articles 3 and 12 of the United Nations Convention on the Rights of the Child (<https://www.unicef.org.uk/what-we-do/un-convention-child-rights/>). The best interests of the children need to be kept paramount. For example, if you intend to collect data from students, teachers can be asked to ensure that students are fully informed of the data evaluation before any data are collected. Children who are capable of expressing their views can be encouraged to do so, commensurate with age and maturity, otherwise the consent of parents can be sought.

It is also important to safeguard the privacy of human subjects and ensure that their participation in the project does not place them in any personal jeopardy because of what they do or say. It should not be possible to attribute opinions, perceptions, or behaviours to individuals or small groups. Confidentiality should be preserved by reporting aggregated data and by placing limitations on the size of the group for which data will be presented in any disaggregated form.

In order to be able to link data collected, different questionnaires completed by one person (e.g. questionnaires before a PD course (“pre-questionnaire”) and after a PD course (“post-questionnaire”), every person (teacher and student) needs to be asked to use an unambiguous code which does not allow identification of the person.

Taking pictures or even filming for the purpose of data collection in classrooms leads to even more ethical issues. Again, information sheets need to be distributed and all parents must sign the consent form. If some parents (or children) do not give their informed consent, you need to arrange the filming in such a way that these students cannot be seen on the film (e.g. by placing them behind the camera).

For more information on ethical issues, please see

- Information on the protection of fundamental rights and freedoms of natural persons: the Universal declaration of Human Rights, (see: <http://html.knowyourrights2008.org/en/universal-declaration-of-human-rights/universal-declaration.html>)
- The right to privacy of persons, with respect to the processing of personal data: EU Directive 95/46/EC – The data Protection Directive.
- See also the Data Protection Act (1998) in the UK (see: http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1), the Data protection act 2002, (see: http://docs.justice.gov.mt/lom/legislation/english/leg/vol_13/chapt440.pdf) in Malta or the Hungarian data protection law (see: http://abiweb.obh.hu/dpc/index.php?menu=gyoker/relevant/national/1992_LXIII).

1.3 A possible model for evaluating PD programmes

It is not uncommon that PD initiatives are described as “successful” without a clear definition of what “successful” PD means. It is important to clarify the aims and learning objectives of a PD programme and let these aims constitute the prime basis for decisions about what to evaluate and which methods to use.

A powerful model for reflections on which evaluation methods to use in relation to aims and objectives is the five level model of Guskey (2000):

1. **Participants’ reactions** to the PD programme
2. **Participants’ learning** in relation to new knowledge, skills, and attitudes
3. **Organizational support and change**: Here the focus shifts from participants to organizational dimensions that may be vital to the impact of the PD programme
4. **Participants’ use of new knowledge and skills in their teaching practice**
5. **Students’ learning outcomes**: What was the impact on students?

Often, participants of a PD activity are asked to complete a so-called satisfaction questionnaire. Here the participants can for example be asked whether or not they approve of the PD, how much they liked the facilitator and to what extent they consider the topic to be relevant for their day-to-day teaching. These questionnaires and surveys typically include a combination of rating-scaled items and open-ended response questions that allow participants to provide more personalized comments. So, was a PD course successful if many participants were satisfied? We could argue that this is not necessarily the case, although the satisfaction of participants might be considered as a prerequisite for effects at other levels. Such questionnaires evidently only measure the impact at level 1.

Certainly, effects at one or more of all five levels are what we strive for when providing PD. But providing powerful PD can be a challenge: It is not always easy to attend to all participants’ needs and framework conditions at schools may not be supportive. Therefore, it is possible that a PD programme may not have an impact on all participants at the levels considered. Naturally, there will be PD programmes that have more impact on participants than others and are thus more successful. Therefore, the impact of the PD programme at different levels can be seen as an indicator of the quality of the PD programme. For more information about the quality of PD programmes, please see the “Ready-to-use Guide for High Quality STEM Professional Development” (<http://stem-pd-net.eu/en/ready-to-use-guide/>).

A vital question in this respect is how we can evaluate the impact at the different levels. Often providers use questionnaires asking for the teachers’ self-reported views on for example, beliefs practice and/or knowledge. However, again, this does not say much about what goes on in the classroom and

how students profit from it. Of course, there are other methods like classroom observations, interviews with students, analyzing students' learning diaries and quantitative questionnaires or test-based evaluations among students. These methods provide in-depth insight but may make high demands in terms of personnel costs and time invested.

However, Guskey (2014) claims that good evaluations do not have to be costly or complicated. "What they require is thoughtful planning, the ability to ask good questions, and a basic understanding of how to collect appropriate evidence in order to find valid answers. In many ways, good evaluations are merely the refinement of everyday thinking. They provide sound, meaningful, and sufficiently reliable information that allows thoughtful and responsible decisions to be made about professional learning processes and effects."

1.4 Overview of evaluation methods

When considering what evaluation method to use, this is of course related to

- the objectives and the indicators you intend to measure.
- the level of impact you refer to (as outlined in section 1.3).
- the contextual conditions, such as time and money available for evaluation as well as the willingness of teachers, students or schools to participate in the evaluation.

There are several methods for evaluation. You can administer questionnaires, do interviews or observations or carry out a desktop analysis of already existing data, such as school reports or minutes. Naturally which level of the Guskey classification they address is also connected to the target group the method is applied to. An interview with a teacher will lead to different insights (mainly on level 1 and 2) than an interview with the head of school (mainly level 3). Questionnaires completed by teachers will give information mainly on level 1 and 2, questionnaires completed by students mainly on level 5 and from the students' perspective on level 4.

Therefore, in the following (see table 1 & 2) we will give an overview of the evaluation methods presented in this guide in relation to the target group they are applied to. We will also indicate to which level of the Guskey levels they refer to. The table shows, that there is no one and only way to evaluating a certain level of impact. There are always several methods to select from, and of course different methods can be combined. For example, you could combine classroom observation with the analysis of a lesson plan. If questionnaires asking about how lessons were perceived are completed by teachers and students, this might give you insights into the way lessons were held.

Table 1 gives an overview about more typical evaluation methods. They are described in part 2.1. In Table 2 the more innovative methods are listed. You will find them in section 2.2.

	Method	Level Participants' reaction	Participants' learning	Organization support	Participants' practice	Students' learning outcomes
1	Satisfaction questionnaire	X				
2	Teachers' questionnaire	X	x	x		
3	Semi-structured interviews for participants	X	x	x	x (as perceived by teachers)	
4	Learning Diaries	x	x		X	
5	Classroom observations				X	
6	Students' questionnaires or interviews				x (as perceived by students)	x

Table 1: Overview of standard evaluation methods

	Method	Level Participants' reaction	Participants' learning	Organization support	Participants' practice	Students' learning outcomes
1	Bringing the best three tasks				x	
2	Presentation of lessons				X	
3	Response Systems	x	x	x		
4	SWOT-Analysis	X	x	x	x	

Table 2: Overview of innovative evaluation methods

1.5 Designing the evaluation concept

As outlined above the first step to designing your evaluation is to clarify the aims and objectives of the PD programme and define measurable indicators. The next step would be to decide on which of the five levels of Guskey (2000) the impact should be evaluated (see 1.2) (whilst also taking into account ethical issues).

The next decision you need to make is if you want to carry out a summative or a formative evaluation:

- **Summative evaluation** refers to assessing the outcome of a program. It means evaluating participants' development at the end of a PD programme. Typically this status is compared with the status at the beginning, a standard or a benchmark. In a PD programme this is typically done by asking teachers (or even students) to complete a pre-test (before the start of the PD course) and a post-test (after the end of the PD course) and compare the results. These tests mainly mean rating scale-items, which can be evaluated by standardized statistical evaluation tests (so-called *quantitative data*). However, it can be also done by carrying out (retrospective) interviews or classroom observations at the end of a PD programme. Interviews or classroom observations normally produce large amounts of data which can be evaluated e. g. by coding (assigning codes to paragraphs) or by interpretation of texts.
- **Formative evaluation** refers to a range of formal and informal evaluation procedures conducted in the PD process in order to get deeper insights into teachers' development over the course of time: What competencies did the teacher develop at first, which later? How did their perception of challenges change during the course of time? How did their attitudes change in the course of time? It typically involves qualitative evaluation methods, such as semi-structured interviews or classroom observations but also quantitative data can be collected for formative evaluation. In order to give information on the development process data needs to be collected at several points of the PD programme.

The decision on which evaluation to use also depends on the structure of PD programme to be evaluated. For a one-off workshop of three hours it is meaningful to evaluate the satisfaction of participants at the end of the programme. For a summative evaluation of the programmes' output a longer process over several days may be needed. For evaluating the development of teachers and maybe even their teaching formatively, even longer evaluation programmes might be necessary.

2 Evaluation methods

In the following, we will discuss evaluation methods in more detail. We will structure them according to Guskey’s five levels, notwithstanding that there is no unambiguous structure. We will sequence them primarily according to the first level they can address. At the beginning of each method a brief overview of the features of the methods will be given.

In section 2.1 we present “more standard” evaluation methods from the social sciences, which are often described in specialists’ books for data collection and which are also accessible to rigorous data evaluation. In section 2.2 we will describe non-standard methods which might be considered to be more feasible for PD course leaders but are not easily accessible for rigorous data evaluation.

2.1 Standard methods

2.1.1 Satisfaction questionnaire

Level according to Guskey	Participants’ reactions (Level 1)
Target group	Participating teachers
Duration of the course	Short-term
Data produced	Quantitative and qualitative data (depending on the questions)
Evaluation design	Neither summative nor formative, only giving information about teachers’ reactions at the end of the course.
Data collection	At the end of the course
Number of participants	Large

Table 3: Overview for satisfaction questionnaire

If the objective of the evaluation is to focus on level 1 and to evaluate the satisfaction of the participants, this can be made with a satisfaction questionnaire – sometimes also referred to as a “Happy sheet” – in which the PD programme is reflected on. The intention of the satisfaction questionnaire is to evaluate the satisfaction of the participants after a PD course. Typically participants are asked whether they agree or not to statements like:

- The announcement of the course was easy to find.
- The announcement allowed enough time to subscribe for the course.
- The aims of the PD course have been made clear.
- The atmosphere in the course was friendly.
- I can use the materials I got in my day-to-day teaching.
- I learnt useful aspects for my day-to-day teaching.
- The course leader took into account participants’ needs.
- The methods used during the course were appropriate.
- I would recommend the course to my colleagues.

Preferably, agreement to these statements is measured using a four- or five-point scale, ranging for example from “strongly disagree” to “strongly agree”.

In this respect experts often discuss the question whether there should be an odd number of answers to select with a “neutral” – box in the middle. On the one hand, research has shown, that people in this case often tick “neutral” because they are not forced to decide whether they agree or not. This is called an “error of central tendency” (Schnell, Hill & Esser 1999). On the other hand, even numbers of boxes, force a decision, even if the participant really does not tend to go in one direction. In this respect, we need to highlight, that there is no correct answer to this question but that it is important to

reflect on whether odd or even numbers in the rating scale should be used and to have a reason for using this scale (Harvard website)¹.

Such a satisfaction questionnaire gives PD course providers only information on the satisfaction of participants and nothing more. As it stays on level 1 of the impact levels as listed by Guskey (2000), it can indicate whether participants might participate in another PD course, but it will not inform about their knowledge, their beliefs, their teaching and so on.

The satisfaction questionnaire can be also supplemented by so-called “open items”. Here teachers are given space to phrase their opinion in their own words. Such questions could be:

- What did you like in this professional development course?
- What did you not like in this professional development course?
- Do you have suggestions for optimization?
- What do you think is missing?
- What can you use in your class?

These open questions may provide feedback which is not covered by the closed items and has proven to be quite informative.

Depending on the aim of the evaluation, it is also possible to not only look back at the PD activity, but also add some questions regarding the future. Possible items, where participants can rank how strongly they agree, can be:

- The PD activity has an impact on my future teaching.
- I will use what I learned in the future.

This indicates impact on level 2, even though it will not give a deep insight in teachers’ learning. For fully covering level 2, a questionnaire as described in section 2.1.2, would be appropriate.

Tips

- *For a high response rate it is recommended to make the participants fill in the sheet towards the end of the PD programme, but not afterwards.*
- *If evaluating a long-term programme one might want to do satisfaction questionnaires also as it proceeds (so a formative evaluation of the satisfaction). A quick alternative that delivers immediate results is thumbs up and down.*
- *When questionnaires are in electronic format it is recommended that the access to them is easy, including perhaps a QR code.*
- *When questionnaires are on paper it is recommended that they should be given to the teacher before the end of the course.*

Pro	Con
<ul style="list-style-type: none"> • Easy to handle • Direct response • Only little time needed to complete the questionnaire 	<ul style="list-style-type: none"> • Mainly evaluation on level 1 • No information about teachers’ competencies • No insights to the in-classroom teaching of participants or impact on pupils

Table 4: Pros and Cons of satisfaction questionnaire

¹ https://psr.iq.harvard.edu/files/psr/files/PSRQuestionnaireTipSheet_0.pdf?m=1357530492

2.1.2 Teachers' questionnaire

Level according to Gus-key	Participants' reactions (level 1) and competencies (level 2)
Target group	Participating teachers
Duration of the course	Longer term
Data produced	Quantitative and qualitative data (depending on the questions)
Evaluation design	Mainly used for summative purposes, can be also used for formative evaluation
Data collection	Mainly before and after the course (Pre and Post-design)
Number of participants	Large

Table 5: Overview of characteristics of teachers' questionnaire

The most important difference between the teacher questionnaires described here and the satisfaction questionnaires described in 2.1.1 is that teacher questionnaires aim at evaluating impact on level 2, knowledge, skills and attitudes, while satisfaction questionnaires primarily aim at level 1.

In order to evaluate constructs such as skills or attitudes, that might not even be obvious to the teachers themselves, it is not possible to use just one individual question for each point of interest (e.g. beliefs on using inquiry in classroom, their understanding of inquiry, frequency of using inquiry). A number of questions will have to be applied for each point of interest to ensure statistically reliable answers to the questions.

Often, answers are supposed to be given on a 4 or 5-point Likert scale, ranging for example from "I strongly agree" (4) to "I strongly disagree". The question of whether to use a 4 or 5-point Likert scale is again connected to the error of central tendency (see 1.1).

The following example (Figure 1) shows an extract of a questionnaire with statements implying the question "to what extent do you agree or disagree" designed to evaluate the change of teachers' attitudes and beliefs on school mathematics during a PD course on mathematical modelling (applying mathematics to open and complex real world problems). In this case we chose a 5-point Likert scale (Maaß & Gurlitt, 2011) to explicitly allow for the neutral option.

	School mathematics in my lessons from my point of view as a teacher	strongly disagree				strongly agree
5.1.1	School mathematics is a collection of procedures and rules which determine precisely how a task is solved.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.1.2	School mathematics is very important for the students later in life.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.1.3	Central aspects of school mathematics are flawless formalism and formal logic.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.1.4	School mathematics helps to solve daily tasks and problems.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5.1.5	School maths is characterized by rigor – rigor in relation to definitions and a formal rigor of mathematical argumentation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 1: Extract of a questionnaire designed to evaluate the change of teachers' attitudes

The statements need to be formulated in a way that is easy to understand and contain only one question. For example, the sentence "School mathematics is very important for students later in life and for society as such" is not appropriate as a statement, because it contains two aspects. If you use such a "double" item, it is difficult to see, what a certain answer means. For example, if they tick the middle, does it mean that they agree with one statement and not with the other? Or does it mean that they agree partly with both?

As far as possible, only questions from evaluation scales which have already been piloted and used should be applied as they are likely to produce more reliable and valid results (a scale here means a number of items designed to evaluate the same aspect, for example teachers' attitudes about mathematical modelling in class). For example, Maaß & Engeln (2018), adapted three student measurement series of the OECD PISA study (OECD 2009, pp. 333-336, series: student investigation, interaction, focus on models or applications) to evaluate the effects of a PD course on modelling and inquiry-based learning (IBL) in the project Primas (Maass & Doorman, 2013). They did so because the PISA measurement series conformed to their understanding of modelling and IBL (OECD 2016, pp. 69-72) and were tested at an international level and therefore ensured valid and reliable testing in the project Primas. The evaluation design in Primas aimed at collecting data from teachers and students. Therefore they took the student questions of the PISA series as they were. To make them fit for teachers as well, they rephrased them. For example they used for teachers "In my lessons I explain the relevance of this subject in our daily lives" instead of the statement for students "The teacher clearly explains the relevance of science concepts to our lives" (ST34Q15). Table 6 shows the items used in the teacher questionnaire.

Items on teaching and learning practice, teachers' questionnaire.		
Name	Item	Pisa 2006
inv1	In my lessons the students design their own experiments/investigations.	(ST34Q 08)
inv2	In my lessons the students do experiments/investigations to test out their own ideas.	(ST34Q 16)
inv3	In my lessons the students have the chance to choose their own experiments/investigations.	(ST34Q 11)
stc1	In my lessons the students are given opportunities to explain their ideas.	(ST34Q 01)
stc2	In my lessons the students have discussions about the topics.	(ST34Q 13)
stc3	The students are involved in class debate or discussion.	(ST34Q 09)
aut1	In my lessons I use this subject to help the students understand the world outside school.	(ST34Q 12)
aut2	In my lessons I show the students how this subject is relevant to society.	(ST34Q 17)
aut3	In my lessons I explain the relevance of this subject to our daily lives.	(ST34Q 15)

Table 6: Items on teaching and learning practice

It is important to pilot any questionnaire with persons of the targeted group before using it, in particular if you do not use tested items. Piloting helps to minimize unclear questions or questions, which do not have sufficient variance in the answers.

The length of a questionnaire is an issue which needs careful attention. If a questionnaire is too long, the answers in the last part of the questionnaires might be influenced by the tiring-effect on the participants. If the questionnaire is too short, it might not produce reliable results.

If you develop a questionnaire which can be used or at least adapted for the evaluation of all PD courses you run, it is less time consuming, though probably less specific.

A limited number of additional open questions might produce interesting answers, which you might not get with only closed questions. For example, you might like to ask the question “Which aspects discussed in the PD course do you actually use in class?”, “What did you like best in the course and why?”

In order to evaluate the participants’ competencies student tasks can be included in the questionnaire. For example, specific kinds of tasks that the teachers got to know in the PD programme can be given to the teachers and they can be asked to (1) solve it themselves, and then (2) anticipate possible student problems when dealing with the task. Another option is to give the participants students’ solutions to a certain task and ask participants to comment on the quality of the solutions. When including such aspects in a questionnaire, you should pay attention to ensuring that the participants ‘do not feel tested’.

If you use a pre-post design to evaluate the summative outputs of the PD programme and you are able to reconstruct a significant change of attitudes, the question what caused these changes remains open. Hopefully, the PD programme under investigation is the reason for this change but there might be also other reasons for this. For example, a change of the curriculum or some important political discussion may have given impetus to the change. In order to identify the PD programme as the unambiguous reason for the change, you may wish to evaluate a so-called **control group**. The control group has to complete the same tests, but does not receive the treatment (PD programme). The participant group and the control group need to be similar in terms of numbers of teachers in the group, age, sex, teaching experiences and so on so that they allow for comparison. If then, the changes from pre- to post-test only occur in the participant group, the change can most likely be attributed to the PD programme, if the same degree of change is detected in both groups there is reason to doubt the effects of the programme. There are different methods to assign teachers to the participant group and to the control group. One important method is to assign teachers applying for the course randomly to one of these groups and organize the control group as a waiting group, which participates in the PD course later. However, in reality it often turns out to be difficult to ‘win’ teachers for this random assignment.

Tips

- See *Tips of the satisfaction questionnaire*.
- *Teachers can complete the pre-questionnaire directly at the beginning of the PD course and the post-questionnaire right at the end. If the questionnaires are rather long, it is recommended to serve refreshments alongside.*

Pro	Con
<ul style="list-style-type: none"> • Deeper insights on level 2 (knowledge, skills and attitudes) • A Pre-Post-design gives information about changes 	<ul style="list-style-type: none"> • No insights into the classroom teaching or impact to pupils • A lot of time is needed to prepare such a questionnaire and to pilot it.

Table 7: Pros and Cons of teachers' questionnaires

In such a pre-post study, participants usually complete the questionnaire anonymously. In order to compare the pre- with the post-questionnaire and measure changes by participant, the questionnaires need to have a code. You can for example ask participants to write down the first two letters of the first name of their mother and the birth date of the mother without the year (e.g. 0112 for 1st of December) (see part 2.1).

2.1.3 Semi-structured interviews

Level according to Guskey	Participants' reactions (level 1) and competencies (level 2); and - from the teachers' perspective - level 3 and 4
Target group	Participating teachers
Duration of the course	Longer term
Data produced	Qualitative data (depending on the questions)
Evaluation design	Can be used for summative evaluation when carried out at the end of a PD or for formative purposes in between
Data collection	At the end of the course, at the beginning of the course and in between
Number of participants	Small

Table 8: Overview of the characteristics of semi-structured interviews

Semi-structured interviews serve the purpose of reconstructing a person's perspective in an explorative way (Flick 2012). As opposed to unstructured interviews they allow for a certain focus whilst not restricting the teacher to very limited answers as in structured interviews.

Semi-structured interviews follow a pre-formulated interview guide. The questions in this guide should be open enough to capture the teachers' perspective. In order to develop appropriate interview questions, a clear aim for the interview should be formulated first. For example, the following interview guide has been set up to evaluate the impact of a PD course on teachers in relation to inquiry-based learning (IBL) (Maaß, Swan & Aldorf 2017).

Biography

1. Why did you choose to become a teacher? How long have you been teaching? Subjects?

Mathematics and science teaching

2. Which way of teaching do you consider to be most effective? Why?
3. What are the three most important characteristics of an effective educator/teacher?
4. Describe a good lesson, explain why it was/is good.
5. What is important about mathematics and science?
6. When teaching maths/science to your class, what activities occur in your lesson?
7. What are the most important activities for students in your class?
8. What should students learn in mathematics/science education?
9. What teaching materials/tasks have you found to be educationally useful in the classroom? Why?

Inquiry-based learning (IBL)

10. Please comment on weak and strong points of the current task.
11. In relation to the PD course, are there any changes in your repertoire for promoting IBL? Which?
12. What are the main aspects of IBL? Name the key features of IBL.
13. How much experience do you have working with IBL tasks? What's difficult/easy? Give examples.
14. In how far do you consider yourself prepared for teaching IBL?
15. In relation to the implementation of IBL, describe some of your experiences when collaborating with other staff members. Are your colleagues used to IBL?

Students and IBL

16. In your own words, how do students get on with the implementation of IBL in the

classroom? What is easy/difficult for them? In your opinion, do they like IBL tasks? Why?
17. Describe a few ways that you can positively influence and assist your students in relation to their learning behaviour/study habits?

Figure 2: Interview questions for a semi-structured interview: do you motivate students to become active learners in your classroom?

Depending on the aim and the interview questions an interview with teachers may last between 30-60 min (it may of course be longer or shorter, but this is a reasonable estimate). It also takes some time to make appointments with teachers and to drive to their school, where in many cases the interview takes place. Thus, in comparison to a pre-post questionnaire study, the sample of an interview study is rather small.

Interviews take more time than having teachers completing questionnaires. Therefore only smaller numbers of teachers can participate in the evaluation. Normally, you would not do interviews with all teachers participating in a PD programme.

If you select the sample out of a larger group, the criteria for drawing the sample need to be clear. Did you select rather classical cases or extreme cases? Is it your intention to show a large variety of different cases or rather stick to classical or medium cases? (Yin 2012) How did you decide, which cases to take? Did you rely on questionnaires for selecting the cases (e.g. on the answers to the pre-questionnaire?) or on other observations? Or did you ask participants whether they use the new teaching approach or not?

The advantage of doing interviews (in comparison with for example teachers' questionnaires) is that you get deeper insights as teachers will not tick standardized boxes but tell everything which seems important to them and which they find necessary to say.

You can use the interviews either at the end of a PD programme for a retrospective summative analysis or during a longer term PD programme for formative analysis. You may even carry out a pre- and a post-interview and compare data afterwards in a search for changes.

For analysing interviews you need to transcribe them and select appropriate methods for analysis (e.g. Grounded Theory, see Strauss and Corbin, 1998).

Tips

- *Whilst we found it often difficult to attract teachers for completing questionnaires or classroom observations, we never experienced difficulties in finding teachers for interviews.*
- *In order to avoid interviewees answering according to social desirability it might be helpful to use persons not involved in the PD course for carrying out the interviews instead of the PD course leader.*
- *It is important to establish a nice atmosphere during the interview and to start with friendly opening questions to warm up participants.*
- *As using technology is always tricky it is advised to use two independent recording devices.*
- *To ensure good quality of recordings it is advised to make test recordings in the venue concerned.*

Pro	Con
<ul style="list-style-type: none"> • Deeper insights on level 2 (knowledge, competencies and beliefs) • Pre-Post-design possible – information about a development • Insights to “hidden” concepts and attitudes of participants possible • Follow-up-questions looking into the future are possible – possible deeper insights • Opportunity to understand participants point of view better 	<ul style="list-style-type: none"> • Only subjective view of participants • No insights to the in-classroom teaching or impact to pupils • Preparation time – setting up the interview guide and making appointments • Needs time of provider and participants for carrying out • Needs time to evaluate • Small number of participants

Table 9: Pros and Cons of semi-structured interviews

2.1.4 Learning Diaries

Level according to Guskey	Participants’ reactions (level 1), Participants’ learning (level 2), Participants’ use of new knowledge and skills in their teaching practice (level 4)
Target group	Participating teachers
Duration of the course	Long-term
Data produced	Qualitative data
Evaluation design	Formative
Data collection	During the course whenever applicable, at the end of course
Number of participants	Small

Table 10: Overview and characteristics for Learning Diaries

Teacher diaries and logs can be used in the evaluation of teachers’ PD. They are tools to document and reflect on specific lessons and their impact, and provide a “real-time” opportunity for collecting data on classroom practices (Glennie, Charles, & Rice, 2017).

Researcher-driven diaries (solicited diaries) are a form of diary that individuals are requested to complete (Kenten, 2010). They are often designed and used for research purposes and therefore tailored to elicit specific information, and increasingly popular as a social science tool to elicit in-depth reflections and longitudinal understandings of respondents’ opinions and circumstances. (Filep, Turner, Eidse, Thompson-Fawcett, & Fitzsimons, 2018).

The format for collecting data in diaries varies in several ways. The prompts for data entry are more or less structured, from specific questions requiring teachers to choose or write a short answer (van Meerkerk, 2017) to simple directions regarding a theme to reflect upon and write about during a set period (Filep et al., 2018).

The timing of data entries also varies from regulations of when to make a diary entry (e.g. at certain times during a working day) to instructions to add entries whenever (and almost immediately) anything worth noticing occurs. Diaries can be digital or on paper.

Furthermore, teachers’ reflections on what is influencing their change in thinking and how these changes influence their actions, can be collected in a reflexive log. Diaries and reflexive logs can also be part of a portfolio where they are supplemented by participants’ collections of material that reflects their perspectives and observation, for example photos of situations, classroom films and selected student work. Diaries can be designed and used in many ways in research and evaluation of teachers’ PD (see e.g. Symon, 2004).

Van Meerkerk (2017) discusses a number of instances of using diaries as a method in research and evaluation on teachers' PD. The use of diaries can first of all give room for the "voice" of the teacher, a voice which is seldom heard in research on teaching and learning. Solicited logbooks or diaries can "make public" what would have remained private using other methods, but without being obtrusive, and allow researchers to gather observations from situations where they cannot be present. They allow a study of thoughts, feelings, and behaviours within the natural work context as well as characteristics of the work situation (Ohly, Sonnentag, Niessen, & Zapf, 2010). Furthermore, van Meerkerk (2017) quotes Bolger, Davis, and Rafaeli (2003) who argue that the amount of time between an experience and the account of this experience is minimized using diaries and logs, enabling a dramatic reduction in the likelihood of introspection. Van Meerkerk (2017) also cites research indicating that while biographical narratives or general opinions tend to be grounded in interviews, participants who are asked to keep a record of certain aspects of their lives in diaries tend to be more inclined to mention routine or everyday processes. Compared to questionnaires, diaries and logs can be more useful for collecting information about the sequencing of activities and time spent on different activities by different participants. Keeping a diary or log in which participants record their thoughts and feelings may also help participants to understand their own learning processes. This applies not only to students, but also to the staff involved, who may learn about their own practice.

In pursuing diaries as a research method it is also necessary to deal with some challenges and constraints. All methods for data collection in education can potentially influence the learning process the researcher wants to describe, and asking teachers to reflect on their learning process in a diary is possibly more influential in this respect than other methods. This challenge can be countered by also using other methods and triangulation. Van Meerkerk (2017) argues, with reference to Meth (2003), that solicited logbooks have to be complemented by other methods such as interviews to counter the challenge of decontextualization caused by the subjectivity of the logbook. However, teacher diaries effects on the learning process can also be seen as a benefit, and a tool supporting the aim of the process. Another pitfall discussed by van Meerkerk (2017) is the amount of time and effort required from the participants in sustainable recording of reflections in a diary. It is important that the entries are written soon after the event of interest, and time-related stress can stop teachers from keeping a meaningful diary. Even though each data entry should require little of the data collector's time, the method is resource intensive for the participants involved in the evaluation, and incentives may be needed to encourage participation. Participants may lose motivation in evaluations that require participation over a long period of time, resulting in high drop-out rates, or less informative or less accurate data towards the end of the process.

Van Meerkerk (2017) also describes the use of teacher log books in a PD programme aiming at the development of new skills and attitudes toward art teaching in elementary teachers. The participating schools followed different trajectories with some teachers receiving on the job training, some engaging in workshop sessions taking place in school or elsewhere, and some developing their own lesson series with their team and professional art teachers. Each teacher was asked to keep a log before and after every event connected to the programme. A protocol was drafted for the logs, which was kept as simple as possible for the benefit of the teachers. Participants were encouraged to go beyond the protocol and write additional contributions whenever they wanted.

<i>Beforehand</i>	<i>Afterwards</i>
How do you look back on the last period?	What have you done, and with who? (You may also include a programme and/or a list of participants.)
What do you expect form the next activity?	What was your role in relation to the others?
	What did you find difficult about it, and what was easy for you?
	What did you like about it, and what didn't you like?
	What do you plan on doing with what you have learned?

Figure 3: Example of a logbook for a professional development course

The use of diaries and logbooks in research and evaluation of teachers' PD certainly has its challenges, but also many benefits. Diaries can offer insights into aspects of PD that are not accessible through other methods. When implemented wisely, possibly in conjunction with other research methods, they can enrich the data and give the teachers a stronger voice in the evaluation.

Tips

- *Reflect on: What is the purpose of the diary? Depending on the purpose, you should choose the regulations:*
 - *Give specific questions or only give a topic to write about*
 - *Set a certain time for diary entries, a lower limit for the number of entries or leave it to the participants*
 - *Give regulation on entry length*
- *It might be helpful to encourage participants to include pictures, drawings, student work and so on to put emphasis on something, visualize something or as a further explanation*
- *In order to ensure a lasting commitment, participants may need reimbursement for the time spent on writing diaries*
- *It might be helpful to check on the diaries regularly or discuss the diaries with the group.*

Pro	Con
<ul style="list-style-type: none"> • May influence the learning process positively • Offers deeper and more accurate insights into aspects of professional development • Enriches the data and gives the teachers a stronger voice in the evaluation • Allow a study of thoughts, feelings, and behaviours within the natural work context as well as characteristics of the work situation • Allow researchers to gather observations from situations where they cannot be present 	<ul style="list-style-type: none"> • May influence the learning process negatively • Requires some time and effort • May lead to motivation loss and high drop-out rate

Table 11: Pros and Cons of presentations of learning diaries

2.1.5 Classroom observations

Level according to Guskey	Participants' teaching (Level 4)
Target group	Participating teachers and students
Duration of the course	Longer term
Data produced	Qualitative data
Evaluation design	Can be used for summative evaluation when carried out at the end of a professional development or for formative purposes in between
Data collection	At the end of the course and maybe in between
Number of participants	Small

Table 12: Overview of the characteristics of classroom observations


Classroom observations give insight into what goes on in class and to see in which way a teacher adopts (or at least can adopt) a certain teaching approach promoted in a PD course or not. It is necessary to say "whether a teacher can adopt a certain teaching approach" because naturally a lesson observed is always biased by the fact that it is observed and is thus a positive selection. Still it offers valuable insights in what a teacher can do. Reactions of students in class to such teaching approaches in observed lessons also quite often tell you whether they are used to such approaches or whether it is an unusual lesson to them. Thus, classroom observations are a way to evaluate effects at level 4 of the Guskey (2000) categories.

An option to document the classroom observation is to film it. Filming a lesson is not a trivial task at all, because you need to reflect thoroughly where to install the cameras. Does every group's table need a camera to see what is going on during group work? Should there be a camera focussing on the teacher? But then, this camera should not be fixed as the teacher may change their position in the class. Often the recording of the voice with the help of the camera is not sufficient and you need to install other devices for audio-recording. All this should be well-thought-through and tested before using it in the classroom observation. When filming lessons, ethical issues have to be considered (see part 2.2).

An alternative to the complex filming of a lesson is using an observation sheet. However, in this case observers need to be carefully trained in what to observe and how to take notes of the observations.

Different observers may make different observations and you might not be able to compare them. As opposed to a film you cannot take a second look at the lesson. Nevertheless, classroom observation with an observation sheet is a powerful method of evaluation. Depending on the aim of the evaluation, the determining factors and ethical issues, one should choose to videotape the lesson or observe it with an observation sheet.

Below, you find the extract of an observation grid Figure 4 that was used in the project “MaSDiV – Supporting Mathematics and Science teachers in addressing diversity and promoting fundamental values” (www.mascil-project.eu). It is supposed to structure classroom observations in relation to inquiry-based learning, real-life contexts and diversity.



MaSDiV
Supporting Mathematics and Science Teachers in
addressing Diversity and promoting fundamental Values

Lesson observation grid

Part 1: Setting
Ideally, this part will be completed before the lesson starts or at the very beginning of the lesson.

Name of observer:

Rule of the observer in the MasDiV project:

Date of Observation:

Subject:

Topic:

Grade of observation:

Number of students:

Please describe the setting to capture the context of the lesson and the social environment. (observable diversity, classroom equipment, arrangement of students in the classroom)

Part 2: Inquiry Based learning and use of contexts

During the lesson take notes using the timeline template.

Please take observation notes focussing on how *IBL is used to address diversity and how contexts are used in this lesson*. Be descriptive. Write down quotes to illustrate important incidents.

With every note you take please hint if you observed teacher action (T) or student activity (S)

Time	T / S	Observation	Comment

Part 3: Observers interpretation

Please complete part 3 at the end of the lesson or shortly after the lesson.
Please refer to concrete observations you made during the lesson to support your interpretations.

How was IBL used to address diversity?

Figure 4: Extract of an observation sheet

To get more insight into teachers' intentions and into their competencies in planning a lesson, a classroom observation can be combined with collecting the lesson plan of the teacher and an interview with the teacher after the observed lessons, asking them about their intentions, the design of the lessons and why they reacted in a certain way.

Tips

- Practice filming and audio recording of classes in advance – sometimes it is necessary to audio record in parallel with filming, to understand what is being said.
- You need the ethical permission of all parents and students involved. If you do not get the permission for one or two students you need to reflect carefully on where to position these during the filming so that they are not seen on the film.
- Plan in advance which additional corroborative materials you want to collect, such as pictures, students' worksheets or students' solutions.
- Use of the additional materials should also be permitted explicitly by the participants and – if the participants are under age – by their parents.

Pro	Con
<ul style="list-style-type: none"> • Only way to get insights into the classroom and teacher performance • Insights on the impact on pupils possible • Pre-post-design possible, maybe a change is observable 	<ul style="list-style-type: none"> • Not a trivial task • Time consumption is high • Needs a lot of technical preparation and material • There may be problems on the question of personal protection data of participants and pupils • It is not easy to evaluate the data

Table 13: Pros and Cons of presentations of classrooms observations

2.1.6 Students' questionnaires or interviews

Level according to Gus-key	Students' learning outcomes and students perception of the lessons (Level 5)
Target group	Participating students (and in combination with teachers' questionnaires some information on the classroom teaching)
Duration of the course	Longer term
Data produced	Quantitative and qualitative data possible (depending on the questions)
Evaluation design	Can be used for summative evaluation when carried out at the end of a professional development programme or for formative purposes in between
Data collection	At the beginning and the end of the course and maybe in between
Number of participants	Large

Table 14: Overview of characteristics of students' questionnaires or interviews

Students' questionnaires can serve different purposes. They can give insight into students' learning outcomes (on the level of skills and competences or beliefs) and they can give insight on classroom teaching.

Insight on classroom teaching

Collecting data to measure changes in teachers' classroom teaching is challenging (see 1.4). Teachers' answers to questionnaires may give information on their intentions but not necessarily on the changes actually made. Classroom observations give better insights, but are very time-consuming so that only few cases can be studied (Perrin-Glorian et al. 2008, see also 3.1.5). An alternative method used to get insight in classroom teaching is the combination of teachers' and students' questionnaires (Baumert et al. 2004).

Students can be regarded as experts on different ways of teaching (Clausen 2002; De Jong and Westerhof 2001) as they are exposed to a variety of teachers in different subjects over an extended period of time. This has been documented by several studies, in which students' perspective has been typically aggregated to class means (Baumert et al. 2004; Kunter and Baumert 2006). The perception of students seems in particular valid for the description of daily routines of teaching and social features of teaching, whilst they have a limited validity for the (intended) instructional approach of the lesson (Baumert et al. 2004).

As opposed to students, teachers can be considered as experts on various instructional approaches, methods and lesson features due to their education and teaching experience (Kunter and Baumert 2006). This is in particular true for their own educational intentions, whilst they lack the possibility to compare their teaching to those of other teachers (Baumert et al. 2004).

Consequently, students' and teachers' perception of teaching can refer to different aspects. Whilst both perspectives provide valuable insight into teaching, the perceptions do not necessarily coincide. In this respect, overlaps in perceptions are an important indicator to what extent an instructional approach, which a teacher intends to follow, is visible in daily teaching routines for students (Baumert et al. 2004). In a study on PD, these overlaps are an important indicator for the impact of the CPD on participants' teaching. Figure 4 shows an excerpt of a students' questionnaire on inquiry-based learning used within the project MaSDiV.

4. Thinking about the lessons with your teacher: to what extent do you agree with the following statements?

	During class...	Strongly disagree	Disagree	Agree	Strongly agree
1)	...I have the opportunity to deeply deal with interesting tasks or contents.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2)	...I have the opportunity to explore new topics by myself.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3)	...my achievements get recognized.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4)	...I feel valued.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5)	...I am encouraged to work independently.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6)	...I have the feeling to belong there.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7)	...I can choose by myself how I want to work.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8)	...I have the feeling that other students would help me, if necessary.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9)	...I feel understood by other students.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 5: Questionnaire to evaluate students' perspective on a lesson from the project MaSDiV

Tips

- If the students to be evaluated are still quite young you need to make sure that you phrase the questions in a way, that students understand them.
- If the students are even younger or have difficulties in reading, you may need to read the questions to the students when they are completing the questionnaire.
- The students of the participants in the PD programme might be different ages, say, age 10 and age 16, and they might therefore need different questionnaires.
- The questions should be posed in a way, that there are no defamatory answers possible.

Students' learning outcomes

Students generally have little problems in completing tests asking for their knowledge and competence as students are used to getting tested. Therefore, it could also make sense to include a part of a test in a questionnaire to evaluate students' learning outcomes.

As the PISA and other big studies show, it is possible to also test competences in dealing with relatively open tasks (e.g. in inquiry task or modelling tasks) by using closed questions. The following question for such a questionnaire was developed within the project Stratum. In Stratum, classroom materials for doing mathematical modelling in class with low achieving students age 11-12 were developed and effects on students and teachers were evaluated. The following question was one of the items developed to test students modelling competence (Mischo & Maass 2013):

Linda lives in a tower block. She wants to work out how many people live in the same building as her. The block has 8 floors. On every floor there are 4 flats. Every flat has one kitchen, 1 bathroom, 1 living room and 3 bedrooms. How many people approximately live in this building?

b) Which of the following facts could be important for finding the solution to the problem? (You can tick more than one answer!)

The house is 30 meters high.

Roughly 4 people live in each apartment.

There are 2 lifts.

Each lift can hold up to 6 people.

There is one cellar for each apartment.

c) Which calculation could lead to finding an approximate solution to the question? (You can tick more than one answer!)

- O $8 \times 4 \times 4$
- O 6×30
- O $30 + 1$
- O $4 + 2 + 6$
- O $8 \times 4 \times 5$

Naturally, such tests need to be adapted to the age of the students.

Figure 6: Evaluating students' learning outcomes on modelling in the project stratum

Tips

- *It is necessary to test the questions and answer prior to their use with students of the same age group in order to see whether the questions test the full range of competences.*
- *Make sure that students have enough space on the questionnaire for calculations, sketches etc.*

Pro	Con
<ul style="list-style-type: none"> • Insights into the classroom and teacher performance • Insights to the impact on pupils • Pre-post-design is possible, maybe a change is observable • Comparing pupil answers and teacher answers possible 	<ul style="list-style-type: none"> • Subjective views of pupils • Quality of answers depends on age of the students • Preparation time – selecting and formulating items (reliable and valid) • Tailor the items to the age of the pupils

Table 15: Pros and Cons of presentations of students' questionnaires or interviews

2.2 Innovative methods

As announced previously, we will describe non-standard methods in this section, which might be considered to be more feasible for PD course leaders but are not easily accessible for rigorous data evaluation.

2.2.1 Bringing the best three tasks

Level according to Guskey	Participants' teaching (Level 4)
Target group	Participating teachers
Duration of the course	Can be used in short term PD, but is especially useful in longer term PD
Data produced	Qualitative data
Evaluation design	Can be used for summative evaluation when carried out at the end of a PD or for formative purposes in between
Data collection	At the beginning of the course, in between or maybe at the end of the course
Number of participants	Medium

Table 16: Overview of characteristics of semi-structured interviews

A method which requires less effort in collecting information about classroom teaching than video studies or classroom observations is to ask teachers to bring three tasks they used in their lessons, that fit the PD topic most. If it is a longer term PD activity, one can say, that they should bring tasks, they used in between the PD sessions in their teaching practice. Clearly, this is a positive selection, but it provides good information on what teachers understood and are able to implement from the PD programme. Also ask the participants to bring students' solutions to the task such as posters etc. They could also bring pictures of the work (see e.g. Fig. 3) on the task during the lesson for a better illustration of what happened in the classroom.

If the tasks are supposed to be implemented in class in between two sessions then keep in mind whether or not the time-lag in between is big enough for implementation to take place. Not all lessons are appropriate for implementation. Also keep in mind, that no implementation can take place during holidays.

It is very helpful for participants and the efficiency of the PD program, if a time limit to the presentations in the course is set for the presentations and communicated in advance e.g. in the preceding session. Depending on the PD programme, it might be better to have all participants report on a task, or just a few (voluntary or random), or split up the presentations in several sessions.

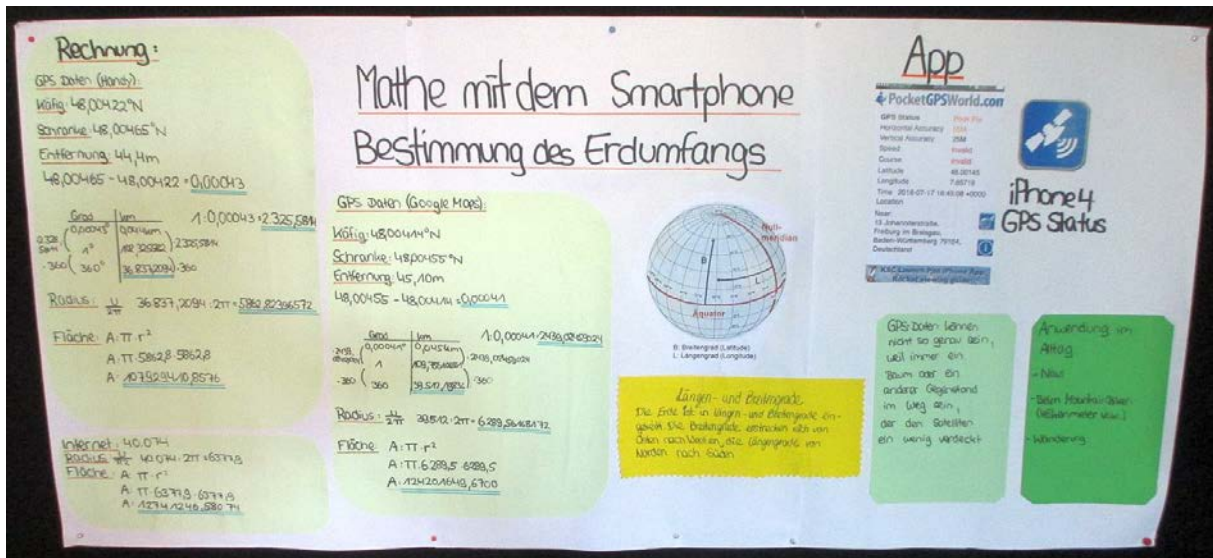


Figure 7: Example for student's poster

Alternatively or additionally you can ask teachers to bring the last two or three written class tests they used in their class. This again will provide interesting information, as it shows how seriously teachers take a certain teaching approach ('What you assess is what you get' Niss 1993).

In the PD course, these tasks can be presented by the teacher and then discussed by the group.

Tips

- Ask the participants to take pictures during the lessons, so that the implementation of the tasks is illustrated in a better way.
- Let the participants bring students solutions for example posters or solutions of tests.
- In a longer term PD programme, the time-lag between the sessions has to be big enough, so that there is time for implementation in between. Only then can reporting on lessons be given as homework to the participants.. Keep in mind: Not every lesson is suitable for implementing the aims of the PD programme. Pay attention to holidays that limit the time for implementation for the participants.
- Set a time limit for the length of the presentations in your PD course and communicate that in advance of the session.
- Think about whether all participants are supposed to present or just a (random or voluntary) selection.

Pro	Con
<ul style="list-style-type: none"> • Maybe offers insights in participants work in their lessons • Insights on level 2 and 3 possible, maybe limited • Insights on level 4 possible, if pupils solution of the tasks are collected too • Pre-post-design possible • References to real cases 	<ul style="list-style-type: none"> • No real evaluation of the PD-event • Only a positively and subjective selection of examples by the participants • Depends on the engagement of the participants

Table 17: Pros and Cons of bringing the best three tasks

2.2.2 Presentations of lessons in the seminar

Level according to Guskey	Participants' teaching (Level 4)
Target group	Participating teachers and students
Duration of the course	Long-term
Data produced	Qualitative data
Evaluation design	Can be used for summative evaluation when carried out at the end of a professional development programme or for formative purposes in between
Data collection	At the end of the course and maybe in between
Number of participants	Medium

Table 18: Overview of the characteristics of presentations of lessons in the seminar

An even better insight into the participants' teaching practice, than bringing singular tasks is the presentation of one or a series of lessons. This is especially helpful, if the PD programme is long term, and the participants have the opportunity to implement their learning in their teaching practice in between sessions. The presentation of their lessons could then be part of the follow-up sessions. For example, you could ask participants to report on how they introduced mathematical modelling in their classes at the beginning of lower secondary education.

In order to get a good structure to the reports it could be helpful to give the participants a grid they can use for preparing and reporting on the lesson(s), though this depends strongly on the aim of the PD programme. In some cases this could limit the freedom of the participants too much. If a time limit for the presentation and the method of proceeding with the presentations (everybody, a voluntary selection, a random selection) is communicated to the participants in advance, the planning of the PD sessions is easier. The more the participants bring into the PD activity from their lessons, the more insights the other participants get. Therefore pictures, students' work or even short film sequences of the lessons are very helpful. In some cases it is also possible to get students to come to the PD session and report on it. Keep in mind the setting of dates for the PD sessions, so that it allows implementation in the classroom in between. Especially consider holidays, that limit the teaching time and pay attention to the fact, that not all lessons are appropriate for the implementation of the PD programme's goals.

Such a presentation of a lesson, possibly integrating students' solutions, pictures or even film sequences, can be also a good starting point for discussions in the seminar.

Tips

- Prepare a reporting grid for the participants.
- Ask the participants to take pictures, collect students' work or make even film sequences during the lessons, so that the implementation is illustrated better.
- In a longer term PD, the time-lag between the sessions has to be big enough, so that there is time for implementation in between, so that it can be given as homework to the participants to report on lessons. Keep in mind: Not every lesson is suitable for implementing the aims of the PD programme. Pay attention to holidays that limit the time for implementation for the participants.
- Set a time limit for the length of the presentations and communicate that in advance of the session.
- Think about whether all participants are supposed to present or just a (random or voluntary) selection.

Pro	Con
<ul style="list-style-type: none"> • Offers insights in participants work in their lessons • Insights on level 2 and 4 possible, maybe limited • Insights on level 5 • Pre-post-design possible • References to real cases 	<ul style="list-style-type: none"> • No real evaluation of the PD-event • Only a positively and subjective selection of examples by the participants • Depends on the engagement of the participants

Table 19: Pros and Cons of presentations of lessons in the seminar

2.2.3 Response Systems/ Digital response systems (as real time evaluation)

Level according to Guskey	Participants' reactions (level 1), Participants' learning (level 2), Organizational support and change (level 3)
Target group	Students, participating teachers
Duration of the course	Long-term or short-term
Data produced	Quantitative
Evaluation design	Mainly: formative (if applied during the course); summative (if applied at the end of the course)
Data collection	During the course whenever applicable; at the end of course
Number of participants	Large

Table 20: Overview for Response Systems

Immediate response systems have been used since the beginning of this century to determine in real time whether or not a course is to the benefit of the participants. Likewise, it can also be used by the teachers themselves in their classrooms as a method of immediate evaluation.

Clickers² are relatively expensive wireless operated devices but have achieved a high level of popularity in PD centers and higher education. Other technology such as the use of mobile phones or laptops (Blicker, Socrative or Kahoot), or Plickers, where the lecturer uses one mobile phone and provides cards for the participants, can reduce the costs for the provider of the PD, if other systems are not available.

The use of these immediate response systems increases participants' engagement, improves peer interaction and formative feedback as well as producing a positive effect on learning (Walklet, 2016). Since learning is typically being assessed when collecting a participants' response, this form of evaluation would be level 2 on Guskey's hierarchy of PD evaluation.

Typically PD providers will ask the participants a question and be able to see the results immediately, either privately for themselves or publicly for all participants. If everybody can see the results, then the provider can decide, whether the responses can be seen anonymously or not. When using clickers, they will have to be numbered, in order to get individual responses.

Blicker³ uses Bluetooth and requires participants to use their smartphones or windows 10 devices. Participants download an App, while lecturers download the Teacher App.

If Plicker cards⁴ are being used, different sets of cards will have to be handed out in order to secure the anonymity of the responses. If everybody gets the same set of cards, participants will be able guess somebody else's response by the symbols. The cards can be downloaded from the internet as well as the app for the mobile phone.

When the participants show their cards, the provider of the PD scans the entire class with their phone, as the participants lift up their cards, to obtain an instant summary, on their phone, of how the class "feels" at this point.

The ability to have instant feedback in a way which anonymises the participants has only been available with modern technology. A show of hands fails to recognise the enormous impact of peer pressure to conform. If the summary of the responses is shown to the whole course, then peer pressure is interestingly applied to the now anonymous individuals, who alone know that they were not with the course, should the majority response indicate that most were understanding the topic.

² See <https://www.turningtechnologies.com/>

³ See <http://www.theinteractivestudio.com/blicker/>

⁴ See <https://get.plickers.com/>

Individual data are easy to obtain in this way (except, at the time of writing, on Blicker) without interfering with the flow of the course or delivery. Such data can easily be stored to create participants' profiles. In addition, participating teachers can learn how to use this evaluation method for their own class and students.

As always with technology there is the risk of breakdown, flat batteries in the clickers or mobile phones, graffiti on the plickers etc. This needs to be planned for and should no more be missing from a modern instructor's workshop/lesson plan than the provision of extra content for early finishers. Using mobile phones also requires that school policy will allow their use. Table 21 shows a comparison of the different devices.

Student Response System	Pro	Con
Clickers	Turning Point software allows high degree of interactivity within a power point presentation.	Expensive Batteries need monitoring Web based Text responses are difficult
Plickers	Very cheap. Only the lecturer requires a smartphone.	The cards may be difficult to manage. Closed responses only
Blicker	Blue tooth connectivity. participant App is free, very small cost for lecturer App	Requires smartphones or other windows 10 devices. Text responses are difficult Cannot collect participant data
Socrative and Kahoot	Free apps available	Web based Need smartphone or web connected device.

Table 21: Comparison of device for immediate response systems

Tips:

- Take some time to get familiar with the technology you want to use!
- Take good care of the technology you use: Where can you get it from? Is it working?
- Is there a possibility for technical support in case of any trouble?
- The method should not be used as an end in itself!

Pro	Con
<ul style="list-style-type: none"> • Instant feedback • Can be used in class as a peg for discussions • Anonymous • Individual data is easy to obtain • No interfering with flow of lesson • Data can easily be stored to create student profiles 	<ul style="list-style-type: none"> • Risks of using technology, such as breakdowns or low battery • Technology has to be provided • Allowance of use from the school policy • Needs quite some technical material and preparation

Table 22: Pros and Cons of Response Systems

2.2.4 SWOT Analysis

Level according to Guskey	All levels except students' learning
Target group	Teachers, PD course leaders
Duration of the course	Short-term, Long-term
Data produced	Qualitative data
Evaluation design	Formative
Data collection	During a PD course or within the organisation
Number of participants	Small

Table 23: Overview for SWOT Analysis

Another option for a more innovative evaluation method to evaluate processes and structures of an institution is the SWOT analysis. "SWOT" stands for "Strengths, Weaknesses, Opportunities, Threats".

It can be used as a simple analysis scheme for a self or group analysis of organizations such as PD centres and it can be also used in a PD course to have the course evaluated by participants.

A SWOT analysis is carried out in two steps. First, participants should think of the present situation of the organization evaluated.

(1) *Think of your organization/school as it is at present (situation analysis), e.g.*

- the services that you provide to people (e.g. pupils) and any other interested parties,
- the internal relations in your organization (e.g. school) and the organization of the work,
- the image the public has of your organization,
- other important aspects of the organization.

S	W
O	T

Figure8: Graph of SWOT Analysis

Participants are now asked to divide a piece of paper in four sections, one each for Strengths, Weaknesses, Opportunities and Threats. Now everyone should enter the current strengths and weaknesses of the organization in the corresponding fields of the first line of the SWOT schema (see figure below).

In the second step, the environment of the institution is assessed.

(2) *Think of the social environment of the organization, e.g.*

- educational and regional development,
- social and cultural development,
- demographic development,
- technological development.

Participants are asked to write down, which opportunities and threats result from this for their organization and enter these in the corresponding fields in the second line of the SWOT schema as well (see figure below).

STRENGTH	WEAKNESSES
We like, that... That should remain.	That bothers us. That needs improvement.
OPPORTUNITIES	THREATS
Approaches are there. This should be further developed.	There are threatening developments. Prevention is necessary.

Figure9: Example for SWOT Analysis

By thinking and reflecting about the four areas (“What are our organizations’ strengths, weaknesses, opportunities and threats?”), the actual state of the organization (or a subunit) can be realised. Based on the strengths, further developments of the organization can be planned and implemented. Based on the threats, changes can be made to offset any damage.

Experiences with SWOT Analysis

Investigating existing situations creates a 'questioning attitude' and supposed 'self-evidences' are partially relativized. Discussing and elaborating topics, questions and survey tools increases communication with each other and in some cases leads to a common awareness of what is important in the organization. This phase is often associated with a spirit of optimism that feeds wide expectations of rapid change, which occasionally leads to gruelling 'maximum programs' (extensive surveys of all possible aspects, innovations at many levels). It can also lead to making some subtle climatic changes visible, resulting in particular from intensive communication.

It should be noted, that SWOT Analysis can also be used as a tool to evaluate the PD Course itself. In this case, the PD Course is the “product / organisation” that is being evaluated. Participating teachers are then asked to think of strengths, weaknesses, opportunities and threats of the PD Course and write them down. Again, the opportunities can be used for further improvement of the course. Strengths show what is especially relevant for participants and can be helpful for a scaling up or further dissemination of the courses content. Threats and weaknesses can be helpful to discuss what should be changed or adapted.

Tips:

- *Opportunities should lead to realistic goals. It might be helpful to break down goals to make changes visible.*
- *It might be helpful to work on the two steps (two lines of the SWOT schema) one by one and to even have a group discussion in between, to have participants focus on only two topics at a time*

Pro	Con
<ul style="list-style-type: none"> • Increases communication between team members • May lead to a common awareness of what is important to the organization • Intensive communication may lead to climatic changes • Can lead to common aims, that concentrates the resources and knowledge of all participants • Reveals the structures and processes of an organization and gives opportunities to look at them from different point of views • Discussions on threats may help preventing them 	<ul style="list-style-type: none"> • May lead to high expectations regarding a rapid change of organizations' processes and structures, that cannot be fulfilled in the expected time • Requires participants' commitment and willingness to change some processes and structures

Table 24: Pros and Cons of SWOT Analysis

3 Outlook

The evaluation methods described in section 2.1 are typically used in research to assess the impact of PD courses. They give a valuable insight into the impact of PD courses. However, they might not always be feasible in day-to-day PD practice. The alternative methods provided might be more suitable.

The pragmatic description of the different methods, the pros and cons given as well as the tips might have given valuable information on the evaluation methods. In the end, however, it is the PD provider who will decide which method to use. A researcher aiming at the detailed impact of the course on participants will choose different methods than a PD provider interested in the general contentment of participants and possibly looking for ways to promote institutional change in their PD centre.

As has been outlined before, the decision on the evaluation method also depends on the specific objectives of the PD course and its content, measurable indicators, the level of insight we want to give, the amount of effort we want to spend on evaluation, the design of the course, the duration of the course and the number of participants.

This guide provides guidance for practitioners. If you want to share your experiences when using the guide with us or if you have feedback or comments, or missed aspects, please feel free to contact us: icse@ph-freiburg.de.

References

- Altrichter, H. & Posch, P. (2007). *Lehrerinnen und Lehrer erforschen ihren Unterricht [Teachers Investigating Their Work]*. Bad Heilbrunn: Klinkhardt.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). *Mathematikunterricht aus Sicht der PISA-Schülerinnen und Schüler und ihrer Lehrkräfte*. In P.-K. Deutschland (Ed.), *PISA 2003 - Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (pp. 314-354). Münster: Waxmann.
- Böhm, A. (2002). *Theoretisches Codieren: Textanalyse in der Grounded Theory [Theoretical coding: Analysis of texts in Grounded Theory]*. In U. Flick, E. v. Kardorff & I. Steinke (Eds.), *Qualitative Forschung, Ein Handbuch [Qualitative research, a handbook]* (pp. 475-485). Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag GmbH.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). *Diary Methods: Capturing Life as it is Lived*. *Annual Review of Psychology*, 54(1), 579-616. doi:10.1146/annurev.psych.54.101601.145030
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? [Quality of Instruction: A Matter of Perspective?]*. Münster: Waxmann.
- De Jong, R., & Westerhof, K. J. (2001). *The quality of student ratings of teacher behaviour*. *Learning Environments Research*, 4(1), 51-85.
- Filep, C. V., Turner, S., Eidse, N., Thompson-Fawcett, M., & Fitzsimons, S. (2018). *Advancing rigour in solicited diary research*. *Qualitative Research*, 18(4), 451-470.
- Flick, U. (2012). *Qualitative Sozialforschung – eine Einführung [Qualitative social research – an introduction]*. Reinbek bei Hamburg: Rowohlt.
- Glennie, E. J., Charles, K. J., & Rice, O. N. (2017). *Teacher logs: A tool for gaining a comprehensive understanding of classroom practices*. *Science Educator*, 25(2), 88-96.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks: Corwin.
- Guskey, T. R. (2014). "Evaluating Professional Learning." In S. Billett (Ed.), *International Handbook of Research in Professional and Practice-based Learning*, 1215–1235. New York, NY: Springer International.
- Kaleta, R., Joosten, T., 2007: *Student Resonse Systems: a University of Wisconsin System Study of Clickers*. Educause Centre for applied Research. *Research Bulletin*, vol 2007, Issue 10, May 8, 2007. <http://web2integration.pbworks.com/f/Student+Response+Systems.pdf>
- Kenten, C. (2010). *Narrating oneself: Reflections on the use of solicited diaries with diary interviews*. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(2). <http://dx.doi.org/10.17169/fqs-11.2.1314>
- Kunter, M., & Baumert, J. (2006). *Who is the expert? Construct and criteria validity of student and teacher ratings of instruction*. *Learning Environments Research*, 9(3), 231-251.

Maass, K. & Engeln, K. (2019). Professional development on connections to the world of work in mathematics and science education. *ZDM Mathematics Education*, 51(7). DOI 10.1007/s11858-019-01047-7.

Maass, K. & Engeln, K. (2018). Effects of Scaled-up Professional Development Courses About Inquiry-Based Learning on Teachers. *Journal of Education and Training Studies*, 6 (4), 1-16. DOI: 10.11114/jets.v6i4.3083.

Maass, K. Swan, M., Aldorf, A. (2017). Mathematics Teachers' Beliefs about Inquiry-based Learning after a Professional Development Course - An International Study. *Journal of Education and Training Studies*, 5 (9), 1-17. <https://doi.org/10.11114/jets.v5i9.2556>

Maaß, K., & Doorman, M. (2013). A model for a widespread implementation of inquiry-based learning. *ZDM - The International Journal on Mathematics Education*, 45(6), 887-899.

Mischo, C. & Maaß, K. (2013). The Effect of Teacher Beliefs on Student Competence in Mathematical Modeling – An Intervention Study. *Journal of Education and Training Studies*, 1(1), 19-38.

Maaß, K. & Gurlitt, J. (2011). LEMA – Professional development of teachers in relation to mathematics modelling. In: G. Kaiser, W. Blum, R. Borromeo Ferri, G. Stillman: Trends in the teaching and learning of mathematical modelling – Proceedings of ICTMA14 (S. 629 - 639). New York: Springer.

Mason, J. (2001). *Researching your own practice: The discipline of noticing*. London: Routledge Falmer.

Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.

Meth, P. (2003). Entries and omissions: using solicited diaries in geographical research. *Area*, 35(2), 195-205. doi:doi:10.1111/1475-4762.00263

Niss, Mogens (1993): Assessment of mathematical applications and modelling in mathematics teaching. – In: de Lange, Jan, Huntley, Ian, Keitel, Christine, Niss, Mogens (Hrsg.): *Innovation in maths education by modelling and applications*, Ellis Horwood, Chichester, S. 41-51.

OECD (2009). *Technical Report- PISA 2006*. Paris: OECD Publishing.

OECD (2016). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. PISA: OECD Publishing, Paris.

Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research: An introduction and some practical recommendations. *Journal of Personnel Psychology*, 9(2), 79-93.

Perrin-Glorian, M.-J., Deblois, L., & Robert, A. (2008). Individual practicing mathematics teachers: Studies on Their Professional Growth. In K. Krainer, & T. Wood (Eds.), *Participation in mathematics teacher education. Individuals, Teams, Communities and Networks (Vol. 3, pp. 35-39)*. Rotterdam: Sense Publisher.

Schnell, R., Hill, P., Esser, E. (1999): *Methoden der empirischen Sozialforschung*. 6. Aufl. Oldenbourg Verlag, München 1999, S. 330 f. ISBN 3-486-25043-4 (EA München 1988)

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research*. Thousands Oaks: SAGE Publications.

Symon, G. (2004). Qualitative research diaries. In C. Cassell & G. Symon (Eds.), *Essential Guide to Qualitative Methods in Organizational Research* (pp. 98-113). London: SAGE Publications Ltd.

van Meerkerk, E. (2017). Teacher Logbooks and Professional Development: A Tool for Assessing Transformative Learning Processes. *International Journal of Qualitative Methods*, 16(1).
doi:10.1177/1609406917735255

Walklet, E., Davis, S., Farrelly, D., Muse, K., 2016: The impact of Student Response Systems on the learning experience of undergraduate psychology students. *Psychology Teaching Review* 22(1):pgs 35-48.

Yin, 2013: *Case study research*. Los Angeles: Sage.