Teacher Professional Growth in Summative Assessment and Meaningful Learning: A Case Study in Pre-Vocational Geography Education in The Netherlands

Erik Bijsterbosch
Windesheim University of Applied Sciences
Tine Béneker
Utrecht University
Wilmad Kuiper
Freudenthal Institute of Utrecht University
Joop van der Schee
University Amsterdam

Abstract

Teachers' classroom assessment practices tend to encourage rote learning instead of meaningful learning. To enhance teachers' classroom assessment practices, teacher involvement in assessment construction appears necessary. To foster teacher professional growth in relation to this issue, a professional development programme on summative assessment and meaningful learning in pre-vocational geography education in the Netherlands was designed. In 2016, a prototype of the programme was tested and evaluated in a small-scale case study. The results suggest that the programme was feasible and practical and contributed to change in teachers' knowledge, skills and practices.

Keywords: teacher professional growth; professional development programme; summative assessment; meaningful learning; assessment literacy

The relationship between teachers' classroom assessment practices and students' learning: Introduction

The relationship between teachers' classroom assessment practices and students' learning has received considerable critical attention

(Black & Wiliam, 1998a, 1998b, 2012; Harlen, 2005; Harlen & James, 1997; Black et al., 2010, 2011; Harlen, 2004). A central issue regarding this relationship is that teachers' classroom assessment practices tend to encourage rote learning instead of meaningful learning. Teachers are not always aware of this tendency (Harlen, 2004; Black et al., 2010, 2011), which seems to indicate a discrepancy between teachers' knowledge and beliefs and their practices. Teachers' assessment-related knowledge, beliefs and practices are all part of teachers' assessment literacy. Assessment literacy is a set of competencies including knowledge and skills related to educational assessment and the purposes of assessment (Brookhart, 2011; Xu & Brown, 2016; DeLuca, LaPointe-McEwan, & Luhanga, 2016). These competencies have been translated into several standards for assessment literacy, which are usually supposed to serve as a guide for teachers and teacher trainers (Brookhart, 2011; DeLuca et al., 2016).

To enhance teachers' literacy, the closer involvement of teachers in assessment construction appears to be necessary (Harlen, 2005). When teachers are involved in the construction of test items and the corresponding criteria to judge and mark, this will positively affect not only their practices but also their knowledge, beliefs and values. Teachers who are more involved in assessment construction become more aware of the issue of validity with respect to summative assessment (Black et al., 2010).

To date, however, little attention has been devoted to the assessment literacy of geography teachers in the Netherlands and its relationship with students' learning. Previous research by the authors on internal school-based examinations in pre-vocational geography education in the Netherlands has provided some insight into the knowledge, beliefs, attitudes and practices of geography teachers (Bijsterbosch, Van der Schee, & Kuiper, 2017; Bijsterbosch, Van der Schee, Kuiper & Béneker, 2016). These studies showed that:

- A majority of test items assess a kind of remembering.
- Teachers rely heavily on test items from external sources, such as tests attached to the textbooks, in the construction of internal school-based examinations. This tendency appears to be stronger when teachers are older and have more teaching experience.
- A negative correlation exists between the use of test items from external sources and the estimated percentage of test items that contribute to meaningful learning.
- Teachers appear to overestimate the percentage of test items contributing to meaningful learning.
- Teachers' conceptions of the content and purpose of internal school-based

examinations appear to be highly influenced by high-stakes tests, especially the external end-of-school (exit) examination. The results indicated that teachers use the same formats in their internal school-based examinations as in the external examinations because they believe that these test items give the most reliable results. Therefore, their constructed internal examinations appeared to be characterised by an emphasis on test items that can be reliably marked at the expense of construct validity.

These two studies were part of an overall research design that aimed to support the professional growth of geography teachers in pre-vocational education with regard to summative assessment and meaningful learning. Professional growth refers to a more-than-temporary change in teachers' knowledge, beliefs and practices regarding the relationship between summative assessment and meaningful learning. This intended change is supposed to have a positive effect on teachers' assessment literacy and teachers' practices in earlier years of pre-vocational education and as such contribute to meaningful learning in geography education.

This paper reports on the second phase of the design research: a formative evaluation of a second prototype of a teacher professional development programme (TPDP) on summative assessment and meaningful learning. The research question for this study is:

"How practical and feasible is a teacher professional development programme on internal school-based examinations and meaningful learning in pre-vocational geography education to foster teacher professional growth?"

The aim is to examine and to evaluate to what extent this programme and its components are feasible and practical. First, this study should provide insight regarding which instruction materials, instruments and strategies are considered practical. Second, this pilot study should provide insight into the feasibility of the outline of the whole programme. Third, because this prototype was implemented for the first time in a case study with teachers, this study should also provide insight regarding to what extent the professional growth intended by the programme and the constituting components could be identified.

The outline and the constituting components of the programme, such as instruction materials, instruments and strategies, were implemented and evaluated in a pilot study with six geography teachers in pre-vocational education in the Netherlands in the autumn of 2016. This paper reports on the results of the evaluation of the programme. Before the results are reported, the next section will first give a brief overview of the literature on teacher professional development.

Teacher professional development

To set up a TPDP, several models have been introduced to develop, analyse and stimulate teacher professional development. In 1986, Guskey proposed a model for professional development that was reaffirmed and slightly adjusted in 2002 (Guskey, 1986, 2002). Crucial to this model (figure 1) is that change in teachers' attitudes and beliefs does not come first; rather, it is altered by the successful implementation of new practices and the consequent improved learning outcomes of students. Or, as Guskey stated: 'it is not the professional development per se, but the experience of successful implementation that changes teachers' attitudes and beliefs' (Guskey, 2002, p. 383).
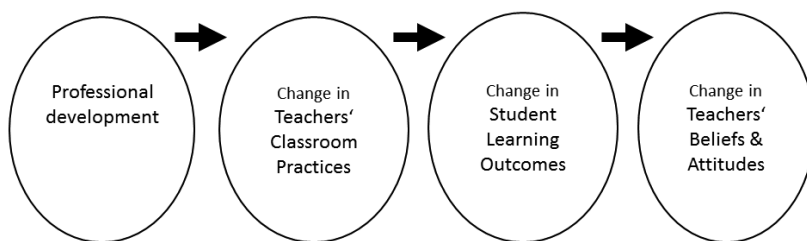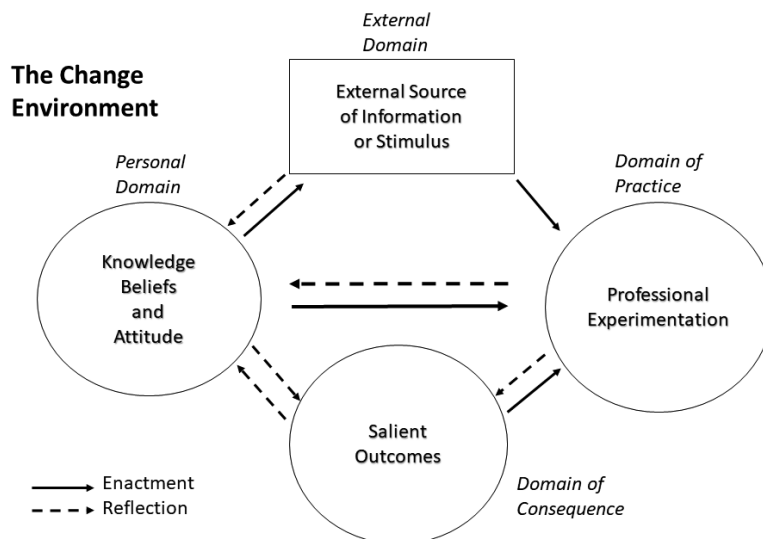

Figure 1 A model of teacher change (Guskey 2002)

Guskey's linear model of teacher change was incorporated by Clarke and Hollingsworth (2002) in their Interconnected Model of Professional Growth (figure 2). This model contains four domains: the personal domain, the domain of practice and the domain of consequences as teacher-related domains one the one hand and the external domain on the other. Unlike Guskey's model, the Interconnected Model of Professional Growth is non-linear. Professional growth can be achieved in multiple growth pathways when lasting changes in and among the personal domain, the domain of practice and the domain of consequence can be fostered. Changes in these domains are initiated by enactment and reflection and are directed by information or stimuli from the external domain. These stimuli from the external domain are considered crucial in directing teacher learning (Voogt et al., 2011) when they focus on curricular enactment (Ball & Cohen, 1996).

Furthermore, it should be noted that professional growth is a complex process that can be accomplished only when teachers learn (Clarke & Hollingsworth, 2002). Teachers learn in the Interconnected Model of Professional Growth as active learners. Active in this sense means that professional growth is not something that is done to teachers but is the outcome of active engagement and reflective participation. Learners shape their professional growth through enactment and

reflection. Enactment in this sense is distinguished from just 'acting', by deliberately translating a belief or pedagogy into action.

The Change Environment

External Domain

External Source of Information or Stimulus

Personal Domain

Knowledge Beliefs and Attitude

Domain of Practice

Professional Experimentation

Salient Outcomes

Domain of Consequence

Enactment
Reflection

Figure 2 The interconnected model of professional growth (Clarke and Hollingsworth 2002, 951)

In addition to this fundamental characteristic of teacher learning, some other characteristics can be identified. First, teacher learning can be accomplished best when the learning takes place in authentic situations (Whitcomb, Borko, & Liston, 2009; Putnam & Borko, 2000), with a focus on subject specific knowledge (Garet et al., 2001). However, the authentic situation is often influenced by contextual factors that might hinder professional growth, such as a perceived lack of time. It is important to acknowledge these factors as much as possible at the beginning of a professional development programme (Clarke & Hollingsworth, 2002).

Second, teacher learning is positively affected by the collaborative design of curriculum materials. Collaboration implies autonomy on decisions about the construction of materials. Collaborative teacher learning becomes more effective when learning is accompanied and scaffolded by a trainer or coach (Bransford, Brown, & Cocking, 1999).

Third, teacher learning in teacher development teams is more effective when it is stretched over time and when teachers have the opportunity to implement intended changes (Bransford et al., 1999; Penuel et al., 2007). To implement intended changes, reflection on intended

outcomes is essential. This reflection becomes more effective when learners reflect not only on their learning goals and learning strategies but also on their beliefs and values (Korthagen, 2004).

Having defined the essentials of teacher development and teacher learning, the final part of this section addresses ways to plan, diagnose and evaluate the professional growth of teachers. To diagnose and evaluate professional growth, it is important first to define the desired outcomes of a TPDP. Subsequently, it is necessary to determine what the training components should be in order to achieve these desired outcomes. To identify the relationship between desired outcomes and training components, an analytical matrix can be helpful (Harland & Kinder, 1997). Joyce and Showers (2002) developed such a matrix (see table 1) to determine how the content of a TPDP can be designed based on desired training outcomes.

Following Joyce and Showers' matrix, the components can be study of theory, demonstration, practice and (peer) coaching. These components of the TPDP are strongly related to the desired outcomes. The percentages in table 1 reflect the percentage of participants likely to attain the desired outcomes when the successive training components are applied. These percentages are an extrapolation made by Joyce and Showers based on research and their experience. Although the estimates are very rough, they give 'rules of thumb for estimating the product of training' (Joyce & Showers, 2002, p. 78). When the desired product of training is a transfer to classroom practices, a TPDP should contain peer coaching in addition to instruction materials (theory), demonstration by a trainer and the collaborative practice of participants.

Table 1. Training components and attainment of outcomes in terms of percentage of participants (Joyce & Showers, 2002)

	OUTCOMES		
Components	Knowledge thorough	Skill strong	Transfer (Executive Implementation)
Study of Theory	10	5	0
Demonstrations	30	20	0
Practice	60	60	5
Peer Coaching	95	95	95

Goals and provisional design principles for the TPDP

A TPDP was designed to foster teacher professional growth in relation to summative assessment and meaningful learning. The goal of this TPDP is to support the professional growth of teachers regarding the construction and scoring of test items in school-based examinations in pre-vocational geography education in the Netherlands that stimulate meaningful learning. To accomplish this goal, the desired training

outcomes of this TPDP are 1) a change in teachers' knowledge and skills, 2) a change in teachers' practices regarding school-based examinations and 3) a change in teachers' beliefs and values.

To attain these desired training outcomes, it is important to know how lasting changes with respect to teachers' knowledge, beliefs and practices can be realised. After all, although the TPDP aims at changes in teachers' knowledge, beliefs and practices in relation to internal school-based examinations, it is also desirable for the programme to affect their conceptions and practices regarding summative assessment in earlier years of pre-vocational education. The ultimate goal is to stimulate meaningful learning in geography education. Changing conceptions and practices regarding summative assessment – and, more specifically, internal school-based examinations – are means to accomplish this goal.

To achieve these aims, the TPDP was set up in line with the Interconnected Model of Professional Growth by Clarke and Hollingsworth (2002). This model incorporates change in teachers' knowledge, beliefs and practices through reflection and enactment, and it is, therefore, highly applicable to designing a TPDP on professional growth with respect to assessment literacy.

Following the aims of the TPDP, the TPDP in this study has the following characteristics:

- The TPDP provides teachers with instruction materials and instruments that enhance teachers' knowledge and skills with respect to the relationship between summative assessment and meaningful learning.
- The TPDP stimulates teachers' core reflection on their beliefs, values and existing knowledge regarding summative assessment and meaningful learning.
- The TPDP stimulates enactment of new knowledge and skills regarding summative assessment and meaningful learning through theory, demonstration, collaborative practice and peer coaching.
- The TPDP contains strategies to stimulate teacher learning through active collaboration.
- The TPDP is situated in teachers' classroom practice but also provides teachers with the opportunity to work collaboratively with colleagues in other contexts outside their own school in order to enable future transfer of knowledge.

Methodology

This study is part of an Educational Design Research (EDR). In the first phase of the EDR, a content analysis on test items in internal

school-based examinations, a questionnaire among geography teachers in pre-vocational education and panel interviews with experts and geography teachers were used to analyse and explore the problem. The outcomes of the first phase, and a literature study on this issue, were used to design the TPDP. In the spring of 2016, a first prototype of the TPDP was evaluated with four experienced teacher educators for expert appraisal. The evaluation was formative and focused on the relevance, consistency and practicality of the first prototype. The outcomes were used to finetune the prototype. The main outcomes of the expert appraisal were:

- The experts agreed that intended professional growth can be accomplished not by a fixed sequence of events but through the interplay of stimuli from the external domain, teachers' beliefs and practices, and students' learning outcomes.
- The experts emphasised the importance of the initial stage of the programme. Teachers should become aware of the goals of the programme and should share the goals of the programme. In addition, the experts suggested devoting more attention in the initial stage of the programme to the problem, i.e., that in school-based examinations, a relatively high percentage of test items assess a kind of reproduction.
- The experts were positive about the sequence of phases in which core reflection on teachers' beliefs was followed by demonstration, collaborative practice and peer feedback.

The amended, second prototype was tested in a pilot study with six geography teachers in pre-vocational education from September to December 2016. All teachers worked in the third grade of pre-vocational education. The first author recruited teacher participants by sending e-mails with an invitation to teachers working in pre-vocational education in the vicinity of the institute of the first author. Teachers were asked to participate in a TPDP on internal school-based examinations and meaningful learning. About 50 teachers were directly invited to participate. Six teachers responded positively to the invitation and participated in this programme.

Outline of the TPDP

In this TPDP, professional growth was pursued by instruction materials from the external domain, which were supposed to stimulate an extension of teachers' knowledge and skills, and by instruments and strategies that were supposed to stimulate enactment and reflection. To accomplish the goals of the TPDP, the programme was executed in three successive phases (see table 2).

Table 2. Outline of the TPDP

Phase I Acknowledgement of pre-existing knowledge and core reflection
1st meeting:
• activation of pre-existing knowledge
• reflection on beliefs and values regarding summative assessment and geography education
• instruction and demonstration on the relationship between test items and learning
Practice and intended enactment (two weeks, at home)
Phase II Extend and internalise knowledge
2nd meeting:
• Demonstration and instruction on test items and cognitive processes of understanding and applying
• Analysis of examples of 'good practices'
• Collaborative practice
• Introduction of scoring rubrics
Practice and intended enactment (three weeks, at home)
3rd meeting: Extend and internalise knowledge
• Demonstration and instruction on test items and cognitive processes, focusing on evaluating and creating
• Analysis of examples of 'good practices'
• Demonstration and instruction on pre-structured test items
• Introduction of strategies as a flow chart for students
Phase III Application in authentic context
Practice and intended enactment, and peer feedback (six weeks, at home)
4th meeting: Evaluation and reflection
• Evaluation and discussion on test items
• Reflection on teachers' knowledge, beliefs and practices

Phase I: Acknowledgement of pre-existing knowledge and core reflection

Before the first meeting was held, the participating teachers were asked to fill in an inventory about their conceptions of the aim of geography education, the purpose of summative assessment in geography education and, more specifically, the purpose of the internal school-based examinations. The outcomes of the inventory were discussed with the teachers in the first meeting, and the teachers were asked to reflect on their deepest beliefs and values related to the outcomes. For this reflection, the phase model of core reflection (Korthagen & Vasalos, 2005) was used.

A second important element of the first meeting was the activation of pre-existing knowledge regarding summative assessment and meaningful learning. This step is essential for teachers to be able to extend their knowledge based on new information. Teachers will either add this new information to their existing knowledge or, in case of pre-existing misconceptions, revise their concepts.

The next step was instruction on how summative assessment could contribute to meaningful learning. To support the instruction, examples of test items were aligned with the learning objectives by a taxon-

omy table based on the revised Bloom's taxonomy (Anderson et al., 2001). This taxonomy table was somewhat adapted to the context of geography education in the Netherlands.

After the first meeting, the teachers had two weeks to practice at home with some of the test items and align them with the objectives. To help them align the objectives with the test items, the taxonomy table was provided as an instrument with the intention to stimulate teachers' enactment. Attention to enactment is vital to accomplish change in teachers' practices (Ball & Cohen, 1996).

Phase II: Extend and internalise knowledge
The next step was to extend and internalise teachers' knowledge. To support teachers in extending and internalising their knowledge, the participating teachers were provided with materials, instruction and demonstration on test items focusing on understanding and applying at the beginning of the second meeting. These materials consisted of some theory with respect to summative assessment and meaningful learning and good practices of test items and scoring rubrics.

Modelling, demonstration and practice were key principles in this second meeting. These key principles were applied in three consecutive steps. First, teachers analysed existing examples of test items and scoring rubrics that were supposed to contribute to meaningful learning. Second, the teachers practised in constructing test items themselves. They could use a flow chart for teachers regarding how to construct these items. Third, these test items were discussed at the end of the second meeting with all participants to attain a mutual understanding with respect to this issue.

In between the second and third meetings, the teachers had three weeks to practise with the test items and scoring rubrics. The teachers were supposed to give each other feedback on the items before the items were discussed at the beginning of the third meeting. Peer feedback was meant to stimulate teachers' self-efficacy by encouraging them through feedback, mastery and vicarious experiences (Bandura, 1989; Schunk, 2003).

In the third meeting, pre-structured test items were introduced as examples of items with the potential to appeal to several cognitive processes, including higher-order thinking processes, in a more structured way for students. Furthermore, the teachers practised with more open and complex test items focusing on cognitive processes such as evaluating and creating. Finally, the teachers received instruments and strategies to scaffold students, such as a flow chart.

Phase III: Apply in an authentic context

In the third and final phase, the participating teachers constructed pairwise test items for their internal school-based examinations. To align the test items with the objectives and the cognitive processes appealing to meaningful learning, the teachers were supposed to use the taxonomy table. The participants were also supposed to provide each other with peer feedback on their self-constructed test items in order to stimulate reflection and enactment. Teachers' reflection and enactment were supposed to be stimulated when they discussed their choices with a peer.

In this phase, the teachers also started to practise with their students. Part of the TPDP were strategies to scaffold students to cope with these items. The teachers used these strategies in their classroom practices to prepare the students for the internal school-based examination. Students' performance in classroom practices was supposed to affect teachers' beliefs towards these test items and consequently their practices.

The constructed test items were discussed with the whole group in the fourth and final meeting. In that meeting, the outline of the TPDP and its constituting components were evaluated with the teachers as well. Before the meeting, the teachers were asked to fill in a survey on the programme and its constituting components. The outcomes of this survey guided the group interview on these issues in the final meeting.
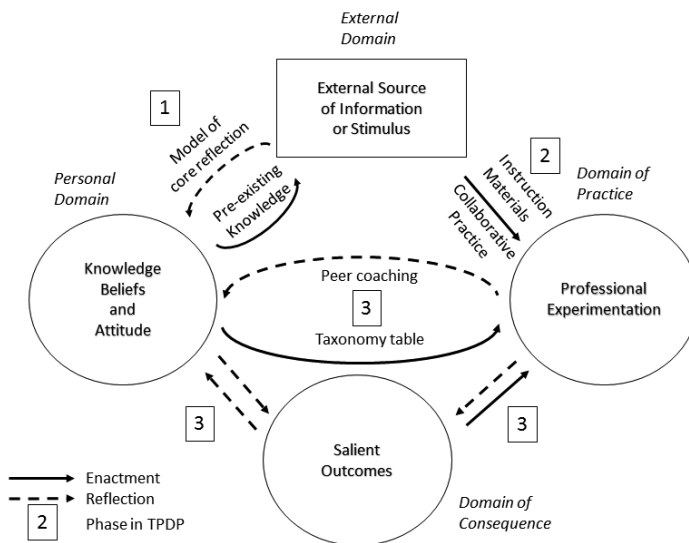
Figure 3 Outline of the TPDP in line with the Interconnected model of Professional Growth

As noted before, the TPDP in this study was set up in line with the Interconnected Model of Professional Growth by Clarke and Hollingsworth (2002). Figure 3 gives an overview of the relationship between this model and the outline of the TPDP in this study. This figure stresses that the TPDP in this study aims not only at change in the three teacher-related domains but also the deliberate stimulation of reflection and enactment.

Data collection

To collect the data to answer the research question, the teachers were asked in a survey about their perception of the practicality of the instruction materials, the instruments, the strategies and the feasibility of the outline of the programme. Materials, instruments, strategies and outlines are considered practical when teachers perceive an element as realistically usable. 'Practicality refers to the extent that users … consider the intervention as clear, usable and cost-effective in 'normal' conditions in the settings for which it has been designed and developed' (Van den Akker, 2010, p. 47).

The teachers filled in the survey anonymously. For each part of the instruction materials, each instrument and each strategy, teachers indicated on a 1-to-5 point Likert scale to what extent the item was practical in relation to the intended outcomes. Second, the teachers were asked why they scored these instruction materials as practical or not practical. At the end of the survey, the teachers were asked to give a score for the whole programme on a 10-point scale and to share their opinion on the programme and the outline of the programme.

The outcomes of the survey were collaboratively discussed and evaluated in a group interview with the participating teachers afterwards. The teachers received an overview of the outcomes before the group interview started. The teachers were asked to reflect together on the outcomes of the survey. The group interview was semi-structured and used the research question as a guideline for the interview. The results from this group interview were also used to answer the research question.

The qualitative data that came from the survey and the group interview were analysed by the first author. Statements of the teachers that referred to the research question were selected. The selection and analysis of the data were approved by another researcher.

Part of the instruction materials was a toolkit with examples of test items, scoring rubrics and strategies for students. An evaluation of the practicality of these geographical test items, scoring rubrics and strategies was reported separately in another paper. The outcomes of this evaluation were only used for triangulation when necessary.

To answer the research question, the following three sub-questions were used to analyse the data from the survey and the group interview:

(1) To what extent are the instruction materials, instruments and strategies in this TPDP practical?

(2) To what extent is the outline of this TPDP feasible?

(3) To what extent could elements of teachers' professional growth, initiated by the instruction materials, instruments or strategies, be identified?

Findings

(1) To what extent are the instruction materials, instruments and strategies in this TPDP practical?

The teachers reported positive perceptions of the instruction materials, especially the examples of test items and scoring rubrics appealing to meaningful learning. The teachers reported in their responses and during the interview less consensus about the theory in the instruction materials. For some teachers, the theoretical background was interesting but time-consuming and less practical. Others mentioned that the theoretical background was practical for them because it helped them become aware of what they were doing. One of the teachers summarised this issue as follows:

"… you are not just practicing, but you also become aware of why you are doing this. And to see lots of examples, that worked for me… You connect (the theory) with your practice. For me, that was a good balance." (Teacher A in group interview)

In the programme, some instruments and strategies were used with the aim to stimulate teachers' reflection and enactment. One of the instruments to stimulate teachers' enactment was a taxonomy table meant to align the objectives for the internal school-based examinations with instruction and testing. The taxonomy table was something the teachers had not used before. The teachers reported that the taxonomy table was practical but complex. One of the teachers reported that the taxonomy table is a practical instrument but not intuitive to apply. The information in the taxonomy table is, according to this teacher, 'overwhelming'. Other teachers mentioned in the group interview that the taxonomy table was helpful to elucidate the objectives.

Another instrument that was introduced to the teachers was a flow chart to construct test items. The teachers reported in the survey that this instrument was practical. One teacher explained in the group interview how he had used the flow chart to align the test items with the objectives for the internal school-based examinations. Although the teachers perceived this instrument as helpful, they also admitted that

the application of the flow chart in the construction of test items made this process time consuming and therefore less practical.

In addition to these instruments, which were meant to stimulate teachers' enactment, some instruments and strategies were used in the programme with the purpose of stimulating teachers' reflection, such as the inventory and model of core reflection at the beginning of the programme. Some teachers argued that these instruments were practical to apply in the programme, while others were less positive. One of the teachers stated in the survey that the purpose of the whole programme could have been made clearer.

The group discussion at the beginning of the first meeting revealed discrepancies between the outcomes of the inventory and what teachers mentioned when reflecting on their deepest beliefs. Teachers' beliefs about the aim of geography education and the purposes of summative assessment seemed not to be perfectly aligned. All teachers agreed or strongly agreed that problem solving is important in geography education, but in the group discussion, they varied strongly in their valuing of higher cognitive processes, such as evaluating and problem solving, as part of summative assessment.

Regarding the questions about teachers' awareness of limitations with regard to their aims, one of the steps in the model of core reflection, the teachers only mentioned limitations that could be labelled as external factors. Most teachers mentioned a lack of time, students' lack of motivation or a lack of equipment, mainly due to limitations caused by a limited number of atlases or inappropriate classrooms. Other striking limitations mentioned were the influence of high-stakes tests, such as the national exit exam, and language as a barrier for students in pre-vocational education. The teachers strongly agreed that the national exit examination has a huge influence on the way they construct their internal school-based examinations. Finally, none of the teachers mentioned limitations that could be labelled as related to their own personal knowledge or beliefs.

A common view amongst the teachers was that the strategy of peer feedback was important. Peer feedback was given on multiple occasions. One such occasion was during practice in between the meetings, and another was when teachers constructed test items for the test. They also gave feedback to one another during the discussion on test items in the final meeting. The teachers highly valued peer feedback and discussion on these occasions with respect to their learning. Teacher B described the practicality of strategies in the survey as follows: "To see each other's test items is very useful, according to me…

To give each other peer feedback is also helpful." Teacher C responded in the survey that "discussion about alignment of test items causes a deepening of your own thinking."

The teachers were less positive about the strategy of the collaborative construction of test items. Some teachers responded that they constructed test items by themselves rather than in collaboration with others. One teacher stated that although he did not construct test items in collaboration with others, this strategy was practical to him.

(2) To what extent is the outline of this TPDP feasible?
The teachers were asked about the extent to which the outline of the programme was feasible. There was a sense of agreement amongst the teachers regarding the feasibility of the programme as a whole. Asked to score the programme on a 10-point scale, all teachers, individually of each other, gave the programme an eight, indicating that the teachers were overall satisfied with the programme. Teacher D explained in the survey this score with the following statement: "I have learned a lot, and not only me but also my students benefit from this."

In the group interview, the teachers shared their opinion on the programme and its constituting components. The teachers highly valued the meetings, especially the second meeting. One of the reasons the teachers mentioned for why they highly valued the meetings was that these meetings gave them a feeling of 'structure'. Another reason mentioned was that these meetings gave them an opportunity to discuss with each other the outcomes of their practice. Therefore, the meetings had, as one teacher mentioned, 'an added value' for the programme.

Although there was less consensus about the constituting components of the programme (see results on sub-question 1), the teachers had only a few suggestions for changing the programme as a whole. A common view amongst the teachers was that they had spent less time on the programme than expected and, therefore, suggested inserting elements to facilitate collaborative work and peer feedback, such as an online community to provide each other with feedback. In the opinion of the teachers, such a community could serve as a 'big stick' to stimulate them to work on the programme at home. The common opinion of the teachers was that peer pressure along with peer feedback would help.

(3) To what extent could elements of teachers' professional growth, initiated by the instruction materials, instruments or strategies, be identified?
To examine to what extent elements of teachers' professional growth

could be identified, teachers' responses from the survey and the group interview were analysed using the Interconnected Model of Professional Growth by Clarke and Hollingsworth. Professional growth refers to lasting changes in at least two of the three teacher-related domains through the mediating processes of reflection and enactment. To examine which elements indicate change in one of the domains or indicate reflection or enactment between the domains, all elicitations were analysed and coded with a scheme that reflected change in a domain or the reflective or enactive links between the domains.

A number of issues were identified. First, change in the personal domain was expressed several times. Change in the personal domain was mainly suggested with respect to teachers' knowledge and skills. Teacher C, describing to what extent their dispositions have changed, commented: "The will to work on construction of test items was already present. The knowledge how to do this was less prevalent."

Change in the personal domain was also reflected by more awareness amongst the teachers of the content and purpose of summative assessment and the relationship with this assessment and students' learning. The teachers' comments below illustrate these changes:

> "I have become more aware what good summative assessments are like and how to construct them." (Teacher D, describing in the survey to what extent their dispositions have changed)

> "I have become more aware to assess what a student knows, but I have also found ways now how to do this." (Teacher B, describing to what extent their dispositions have changed)

A second issue that emerged from the analysis was that elements of change in the domain of practice could be identified. In response to the question about to what extent their tests had changed due to the professional development programme, Teacher D reported: "The whole test has been changed." This statement suggests that this teacher deliberately changed the test, indicating change in the domain of practice. Responses from other teachers also indicated that teachers had changed their summative assessments. The teachers not only reported this change in practice but in some instances also showed that their assessment practices had been changed deliberately, based on external information, thus indicating enactment between the external domain and the domain of practice:

> "I continuously ask myself if students have to relate their pre-existing knowledge with the information in the test item. When they have to do so, this means to me that this is a kind of meaningful learning. The next

step is to categorise the test item (understanding, evaluating etc.)."
(Teacher B, responding in the survey)

Teachers showed not only enactive links between their domain of practice and the external domain but also reflective links between the external domain and the personal domain. Some of their responses indicated that the teachers attached value to the information and materials from the external domain, which appeared to influence their beliefs. Teacher E mentioned: "Most examples in the materials are useful." Statements such as this suggest reflective links between the external domain and the personal domain.

On the other hand, Teacher E also mentioned: "The last example was more difficult because during the classroom discussion with students, confusion arose about the right answer. For some students, this test item was difficult." The teachers' beliefs on the practicality of the examples appeared to be affected by students' difficulties with this example. This statement, therefore, indicates that teachers' beliefs are affected simultaneously through reflective links with both the external domain and the domain of consequence.

A third important issue that emerged from the analysis, therefore, was the relationship between the domain of consequence, i.e., how students perceived and valued this new way of testing, and the personal domain or the domain of practice through reflection. These reflective links appeared to work in two ways. First, students' difficulties with certain test items appeared to decrease teachers' enthusiasm to apply these test items (domain of practice). A common view amongst teachers was that not all test items, especially the ones appealing to evaluating or creating, were applicable and practical for students. One teacher reported, with respect to this issue, that these test items are also more demanding for less literate students.

Second, other responses indicated reflective links between the domain of consequence and the personal domain. Reflection on student performance appeared to stimulate teachers' appreciation of elements from the external domain. Teacher C referred to the strategies students could use in answering the test items: "This (the flow chart) is probably more directing, but it makes it more attainable for students to come up with a good answer."

Other teachers reported that the use of the flow chart helped students structure their answers and, therefore, improved their performance. Feedback to students on the application of this strategy appeared to improve the quality of their answers, as perceived by the teachers. Some

teachers mentioned that they had started to use the flow chart in other classes for this reason. The way students responded to the flow chart, therefore, seemed to have influenced teachers' conceptions positively.

Discussion

In this paper, the aim was to assess to what extent a designed TPDP and its constituting components are feasible and practical to support teacher professional growth regarding summative assessments and meaningful learning in pre-vocational geography education. The design of the programme and instruction materials was based on the outcomes of the analysis and exploration phase of the design study, which showed that teachers hardly construct test items themselves. Second, a content analysis showed that a majority of test items in internal school-based examinations test recalling knowledge. Moreover, when teachers construct test items themselves, they appear to overestimate the percentage of test items appealing to meaningful learning rather than recalling knowledge. The TPDP in this study, therefore, aimed to contribute to a solution to two problems. The first is that in internal school-based geography examinations in pre-vocational education in the Netherlands, a high percentage of test items appeal to recalling knowledge and hardly appeal to the cognitive processes associated with meaningful learning. The second is that teachers hardly construct test items themselves and appear to have problems constructing test items contributing to meaningful learning. Teacher professional development on this problem, therefore, appeared to be necessary.

The results from this study suggest that change in teachers' knowledge, skills and practices through the mediating processes of reflection and enactment, affected by the designed TPDP, could be identified and thus support teacher professional growth. The outcomes of this study suggest that change in the personal domain and the domain of practice is influenced by teachers' interpretation and valuing of student performance.

Teacher professional development in this study was regarded as a non-linear process that depends on the interplay between change in teachers' knowledge, dispositions and practices on the one hand and students' learning outcomes on the other (Clarke & Hollingsworth, 2002). Through reflection and enactment, intended change in one of these domains could foster change in the other domains without a fixed sequence in the TPDP. However, a sequence in the professional growth of the teachers, in line with Guskey's model of professional growth (Guskey, 2002), was apparent. The suggested sequence in this study was a change in teachers' dispositions and practices through the value the teachers attached to student performance.

In their elaboration of the model of professional growth, Clarke and Hollingsworth suggested that a sequence in accordance with Guskey's model could be one of the growth pathways. Nonetheless, according to Clarke and Hollingsworth, teacher professional growth can be accomplished in multiple ways, and 'teacher change often involves multiple and cyclical movements between the analytical domains of the teachers' world' (Clarke & Hollingsworth, 2002, p. 961). Therefore, although change in teachers' knowledge, skills and practices in this study appears to have been influenced by student performance, this does not necessarily indicate that this growth pathway should determine teachers' professional growth in future cases. Rather, it suggests that a future TPDP on the relationship between summative assessment and meaningful learning in pre-vocational geography education should contain multiple cycles and offer participants the opportunity to accomplish professional growth 'consistent with individual inclinations' (Clarke & Hollingsworth, 2002, p. 962).

In line with these implications for a future professional development programme, it seems important to devote more attention to teachers' individual beliefs and values in the initial stage of the programme. Teachers should not only share the goals of the programme but also become aware of the identified problem in relation to their own beliefs and values. In addition, teachers should be given the possibility to find ways to address this problem themselves. When these possibilities are offered, teachers will become more actively engaged in the TPDP, which is an important prerequisite for teacher learning.

In addition to reflection on teachers' deepest beliefs and values, it seems to be important to address teachers' awareness of limitations and their core qualities to realise the ideal or desired situation at the beginning of the programme. Because teachers in this study only mentioned limitations that could be identified as external factors, it might be advisable to pay more attention to teachers' core reflection over a longer period of time. Other instruments, such as a weekly or biweekly logbook, might contribute to more teacher reflection on their limitations and core qualities related to their own knowledge and skills.

Another important outcome of this study suggested that scaffolding students with strategies, such as a flow chart, was important for teachers' change in their personal domain. These strategies were perceived by the teachers to help the students answer test items that appealed to cognitive processes belonging to meaningful learning. Furthermore, by providing these strategies, teachers perceived the quality of the given answers to be higher.

A final important result of this study was that, in general, the participating teachers regarded the outline of the TPDP and its constituting components as feasible and practical. However, according the teachers, there is room for improvement. One of the improvements suggested by the teachers is to use an online community to stimulate exchange of constructed test items and feedback between the teachers instead of pairwise construction of test items and peer feedback. Another suggestion was to use this community in order to facilitate a kind of peer pressure to stimulate teachers to work on the programme. Support by a trainer in the collaborative design process of teams could be helpful (Huizinga et al., 2014).

Conclusion

The current study showed some promising results regarding the professional development of teachers in relation to summative assessment and meaningful learning and the contribution of the programme to this professional development. A note of caution is due here since only six teachers in pre-vocational education participated in this study. The findings may be limited by the small sample size. Furthermore, the selection of participating teachers and students was not fully random. Another source of uncertainty is the type of research applied in this study. The results depended on what teachers reported in the survey and the interviews. Although this method is suitable to explore the reasons for teachers' remarks, more research is needed to verify the results from these groups. The outcomes of this small-scale study will be used to redesign the programme. The redesigned programme will be tested and evaluated with another group of teachers to examine how and to what extent this TPDP fosters teachers' professional growth.

References

Anderson, L. W. et al. (Eds). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives (Complete edition)*. New York: Longman.

Ball, D. L. & Cohen, D. K. (1996). Reform by the book: What is - or might be - the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*. *25*(9), 6-8, 14.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*. *25*(5), 729-735.
https://doi.org/10.1037/0012-1649.25.5.729

Bijsterbosch, H., Van der Schee, J. A. & Kuiper, W. (2017). Meaningful learning and summative assessment in geography education: An analysis in secondary education in the Netherlands. *International Research in Geographical and Environmental Education*. *26*(1), 17-35.
https://doi.org/10.1080/10382046.2016.1217076

Bijsterbosch, H., Van der Schee, J. A., Kuiper, W. & Béneker, T. (2016). Geography teachers' practices towards summative assessments: A study in pre-vocational education in the Netherlands. *Review of International Geographical Education Online*. *6*(2), 118-134.

Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*. *17*(2), 215-232. https://doi.org/10.1080/09695941003696016

Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*. *18*(4), 451-469. https://doi.org/10.1080/0969594X.2011.557020

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*. *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P. & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: NferNelson.

Black, P. & Wiliam, D. (2012). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (pp. 11-32). London: SAGE.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington DC: National Academy Press.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*. *30*(1), 3-12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Clarke, D. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*. *18*(8), 947-967. https://doi.org/10.1016/S0742-051X(02)00053-7

DeLuca, C., LaPointe-McEwan, D. & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*. *28*(3), 251-272. https://doi.org/10.1007/s11092-015-9233-6

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F. & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*. *38*(4), 915-945. https://doi.org/10.3102/00028312038004915

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*. *15*(5), 5-12. https://doi.org/10.3102/0013189X015005005

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*. *8*(3), 381-391. https://doi.org/10.1080/135406002100000512

Harland, J. & Kinder, K. (1997). Teachers' continuing professional development: Framing a model of outcomes. *British. Journal of In-service Education*. *23*(1), 71-84. https://doi.org/10.1080/13674589700200005

Harlen, W. (2004). A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *Curriculum Journal*. *16*(2), 207-223. https://doi.org/10.1080/09585170500136093

Harlen, W. & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*. *4*(3), 365-379. https://doi.org/10.1080/0969594970040304

Huizinga, T., Handelzalts, A., Nieveen, N. & Voogt, J. M. (2014). Teacher involvement in curriculum design: Need for support to enhance teachers' design expertise. *Journal of Curriculum Studies*. *46*(1), 33-57. https://doi.org/10.1080/00220272.2013.834077

Joyce, B. R. & Showers, B. (2002). *Student achievement through staff development*. Alexandria, VA: Association for Supervision and Curriculum Development.

Korthagen, F. A. J. (2004). In search of the essence of a good teacher: Towards a more holistic approach in teacher education. *Teaching and Teacher Education*. *20*(1), 77-97. https://doi.org/10.1016/j.tate.2003.10.002

Korthagen, F. A. J. & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching*. *11*(1), 47-71. https://doi.org/10.1080/1354060042000337093

Penuel, W. R., Fishman, B. J., Yamaguchi, R. & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*. *44*(4), 921-958. https://doi.org/10.3102/0002831207308221

Putnam, R. T. & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*. *29*(1), 4-15. https://doi.org/10.3102/0013189X029001004

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly*. *19*(2), 159-172. https://doi.org/10.1080/10573560308219

Van den Akker, J. (2010). Curriculum design research. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research. Proceedings of the seminar conducted at the East China normal university, Shanghai (PR China)*, (pp. 37-50). Enschede: SLO.

Voogt, J. et al. (2011). Teacher learning in collaborative curriculum design. *Teaching and Teacher Education*. *27*(8), 1235-1244. https://doi.org/10.1016/j.tate.2011.07.003

Whitcomb, J., Borko, H. & Liston, D. (2009). Growing talent: Promising professional development models and practices. *Journal of Teacher Education*. *60*(3), 207-212. https://doi.org/10.1177/0022487109337280

Xu, Y. & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*. *58*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010