University of Groningen

Learning to teach

Tas, Tamar; Houtveen, Thoni; van de Grift, Wim; Willemsen, Melissa

Published in:
Studies in Educational Evaluation

DOI:
[10.1016/j.stueduc.2018.05.005](https://doi.org/10.1016/j.stueduc.2018.05.005)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Tas, T., Houtveen, T., van de Grift, W., & Willemsen, M. (2018). Learning to teach: Effects of classroom observation, assignment of appropriate lesson preparation templates and stage focused feedback. *Studies in Educational Evaluation*, 58, 8-16. <https://doi.org/10.1016/j.stueduc.2018.05.005>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Learning to teach: Effects of classroom observation, assignment of appropriate lesson preparation templates and stage focused feedback

Tamar Tas^{a,*}, Thoni Houtveen^a, Wim van de Grift^b, Melissa Willemssen^a

^a Utrecht University of Applied Sciences, Postbus 14007, 3508 SB, Utrecht, The Netherlands

^b University of Groningen, The Netherlands

ARTICLE INFO

Keywords:

Professional development
Quality of instruction
Lesson preparation and feedback templates
Teacher training college

ABSTRACT

This article focuses on improving the instructional quality of student teachers in elementary education. We developed a coaching approach involving classroom observation and appropriate lesson preparation and feedback templates. Using an untreated control group design with pre-test and posttest ($n = 198$), we answered the question ‘whether student teachers who learned to teach with the new coaching approach achieved a higher level of pedagogical and didactical teaching skills compared to student teachers who did not receive this approach’. The effect variable used was the observation instrument ICALT. We compared the average scores of the control group and the experimental group on the posttest (ANCOVA). The differences found on the posttest, after controlling for the confounding variables, were significant on all ICALT scales. These effect sizes are medium on the scales; ‘Safe climate’, and ‘Clear instruction’, and large on the scales; ‘Classroom management’, ‘Activating pupils’, ‘Adaptive teaching’, and ‘Teaching learning strategies’.

1. Introduction

Results of international comparative studies in pupil results in the Netherlands show a slightly downward tendency (Meelissen & Punter, 2016; Mullis, Martin, Foy, & Drucker, 2012). Against this background, the policy of the Ministry of Education, Culture and Science is focused on increasing pupil achievement. Meta-analysis of the efficacy of the educational system shows that policies affecting the primary processes of learning and instruction are the most successful. The quality of teaching staff, and especially the quality of their instruction turns out to be an important factor in increasing levels of pupils’ achievement (OECD, 2011). As a result, the quality of teacher training colleges is placed high on the national policy agenda (Min. OCW, 2013). In the meantime, several arrangements have already been made to improve the quality of teacher training colleges in the Netherlands. One of them is the development of uniform standards for the teaching profession (10voordeleeraar, 2018; Meijerink, 2012).

In practice, learning to teach takes place within the schools where the student teachers are doing their internship. Elementary school teachers (functioning as mentors) observe their lessons and provide student teachers with feedback. The main guidance instrument during internship is a standard lesson preparation template in which two learning cycles are interwoven: the learning cycle of the children that are being taught and the learning cycle of the student who learns to

teach. Student teachers of all Dutch teacher training colleges have to prepare their lessons on such a mandatory template, and mentors give feedback on the basis of the lesson preparation template and the lesson realization. In the 70 s the so-called DA model ‘didactic analysis model’ (Van Gelder, Oudkerk Pool, Peters, & Sixma, 1971) was introduced in all Dutch teacher training colleges. The model consisted of a didactical cycle from formulating ‘the lesson goal’, ‘the educational needs’ and ‘the actual lesson description’, to the evaluation of the lesson goal’. Van Gelder translated the model into a single template with corresponding observation criteria for the lesson preparation as well as the lesson realization.

In the 80 s and 90 s there was a shift in lesson preparation from the acquisition of teaching skills to reflection skills. The templates based on the ‘didactic analysis model’ did remain practically unchanged, although the corresponding observation criteria were no longer in use and replaced by the student’s own learning goals and (self-) reflection. Up until now, the templates are rarely updated on the basis of recent research on effective teaching skills and effective feedback.

In general the teacher training colleges use one single template during the four training years. These templates do not take the professional development of student teachers into account. Furthermore, the mentors are neither provided with a standardized observation instrument to evaluate the instruction skills of the student teachers, nor with guidelines concerning what kind of feedback would be most

* Corresponding author. Postal address: Hogeschool Utrecht (HU), Postbus 14007, 3508 SB, Utrecht, The Netherlands.
E-mail address: Tamar.Tas@hu.nl (T. Tas).

helpful for the learning process of the student teacher. As a result, the lesson preparation templates do not support the acquisition of teaching skills and the teacher training colleges lack objective data on the teaching skills of their students.

A possible way to address this issue is to improve the quality of the lesson preparation template and to develop a procedure in which the teacher training college and schools work together in collecting objective data on student teaching skills and in which guidelines for feedback are given.

2. Theoretical and empirical framework

For the development of a coaching approach that supports the acquisition of teaching skills of student teachers, we made use of theories on professional development of (student) teachers, the knowledge base on effective teaching behavior, and research on effective feedback.

2.1. Stages in the professional development of student teachers: theory and evidence

In theories on the professional development of teachers, nonlinear as well as linear models are proposed. In the nonlinear models, psychological, sociological and environmental factors are considered major influences on the professional development of teachers (Day & Gu, 2010; Day & Qing, 2007; Fessler & Rice, 2010; Goncalves, 2009; Huberman, 1989; Woods & Lynn, 2001). As a result, the professional development of teachers is supposed to follow a nonlinear path. Since the developmental theories underlying these models start at the induction phase of the teacher, these nonlinear models are less suitable for our purpose.

Several linear models can be distinguished, such as the concern theory (Fuller & Bown, 1975), a five-stage model of the acquisition of expertise (Dreyfus & Dreyfus, 1986), and the life cycle of the career teacher model (Steffy & Wolfe, 2001), where (student) teachers go through developmental phases in a linear process. In the life cycle of the career teacher, the emphasis is on the developmental stages of the teacher after the teacher college training. The stage model of the acquisition of expertise requires as a starting point of the learning process a rather context-free learning environment, which is, in the existing internship conditions at the teacher training college in Utrecht, not practically feasible. Thus, for this study both linear models are also less suitable.

Fuller and Bown's concern theory is the most cited and empirically underpinned theory (Broad & Evans, 2006; Richardson & Placier, 2001). According to the concern theory, student teachers go through three stages during their education. Each stage consists of dominant student concerns – 'something the student thinks about frequently and would like to do something about personally'. In the self-phase, the student teachers have 'concerns' about surviving the classroom. In this stage, students become affected by feelings of inadequacy and dependency. In the following phase, the 'task (of teaching situation) concerns,' ambivalence regarding materials, methods and instruction are of prime importance. Students are able to reproduce the subject matter they have learned, but may be incapable of teaching it. Answering pupils' questions or providing examples remain highly challenging endeavors, and the learning process of the children is yet to become their focus. Nevertheless, in the 'impact phase' ('concerns about pupils'), students demonstrate a greater understanding of their children's learning needs. The time required for a student teacher to progress to a higher stage varies depending on personal characteristics and wider context, and regression to the 'self-stage' remains possible (Fuller & Bown, 1975).

Research conducted with the Teacher Concern Checklist (TCC) shows divergent results regarding a hierarchical sequence of the concerns. Using this TCC, Boz (2008) examines differences in concerns among 339 Turkish students. He observed mainly task concerns, but

witnesses significantly fewer self-related concerns in student teachers' last training year than in the second and third years.

Reeves and Kzelskis (1985) score the concerns of inexperienced students ($n = 128$), compared with those of experienced teachers ($n = 90$) on the TCC. They found no difference in the number of task concerns between the groups; moreover, in both groups, 'impact concerns' were higher than 'self-' and 'task concerns'.

In a number of studies, 'self-' and 'task concerns' have decreased as the students accrued more experience, but 'impact concerns' were occurring from the start.

Pigge and Marso (1997) used the TCC to follow 60 students from the beginning of their education until five years after their appointment. 'Task concerns' developed after 'self-concerns,' also 'impact concerns' occurred, akin to Reeves and Kazelskis' study, from the start. The researchers found that concern changes in the teachers' professional development related to their grade average (students with lower teacher education grades expressed fewer impact concerns) and gender (female students expressed more impact concerns).

Smith and Sanche (1992) found that 'self-concerns' predominated amongst 112 students at the start of an internship above task and impact concerns. By the middle of the internship, task and impact concerns exceeded 'self-concerns'. It is important to note that 'task concerns' started at a high level and remained fairly constant. Such a pattern of overlap of different concerns is corroborated by other researchers.

Watzke (2003, 2007) observed 'impact concerns' in his longitudinal research in first- and second-year student teachers. Seeking possible explanations for the deviations in the linear structure of Fuller's concerns in previous experiences of the students and the specific nature of the studied course, Watzke (2007) argued that the impact phases represented a continuous, holistic concern process amongst beginning teachers, recurring over time.

Ghaith and Shaaban (1999) studied the level of experienced 'self-efficacy' and the concerns of experienced teachers. 'Self-concerns' appeared to decrease with increased teacher experience. Nevertheless, after 15 years of experience, the three categories (including 'impact concerns') had declined. The researchers hypothesized that the internship context (for example, teaching in kindergarten or special education classes, or teaching classes to pupils with behavioral problems) helped to explain the prevalence of certain concerns.

A different kind of criticism led to the discussion, whether the concern theory focus on the person occurs at the expense of the training of professional standards (Buchmann, 1987; Zeichner & Teitelbaum, 1982).

Despite of the criticism above on the hierarchical order of the concern stages, there is enough evidence of the existence of these concerns. The consideration that the concerns may not follow in a hierarchical order does not impact our study. By adding a standardized observation instrument to the templates in construction, we can easily diagnose the accurate professional developmental stage of the student teacher and assign the corresponding lesson template. We therefore decide to develop successive lesson preparation and feedback templates which support the professional development of student teachers, based on this theory.

2.2. Effective Teacher behavior

In addition to theories of stages in the professional development of (student) teachers, we can use research on effective teacher behavior. That effective teaching skills are important factors in achieving pupils' learning gain, is generally acknowledged after more than 50 years of research on this topic (for reviews and research see: Ellis & Worthington, 1994; Hanushek, 2011; Hattie, 2009; Levine & Lezotte, 1990; Levine & Lezotte, 1995; Marzano, 2003; Muijs & Reynolds, 2010; Scheerens, 2015; Scheerens & Bosker, 1997).

For this research, observational teacher behavior, i.e. the behavior

that teachers show in their everyday lessons, is of particular interest. Six categories of effective teacher behavior can be distinguished from the above-mentioned research.

The category ‘Creating a safe and stimulating learning climate’ (1) consists of teaching skills which ensure a climate in which pupils feel safe and where mutual respect prevails. The (student) teacher stimulates learning by acting consistently and supporting the pupils’ self-esteem by providing success experiences. ‘Efficient classroom management’ (2) involves practical matters of organization, such as ensuring an orderly course of the lesson and also didactical teaching skills engaging all pupils during the instruction and motivating them to reach for the best. This category also deals with matters of time management to increase learning time as much as possible. ‘Giving clear instruction’ (3) consist of the successive steps of a direct or explicit instruction. On a short instruction to the classroom as a whole, follows guided practice in which pupils practice their skills under the guidance of the teacher. The instruction starts with the clarification of the lesson objective(s) and the activation of prior knowledge and ends with the evaluation of the lesson objective(s). In the category ‘Activating pupils’ (4), teaching skills are central which specifically increase pupil involvement by using work forms that lead up to active learning. Adapting the instruction and classroom assignments to relevant differences between pupils and thus reach optimal learning gain for all pupils, is the content of the category ‘Adaptive teaching’ (5). The category ‘Teaching learning strategies’ (6) supports the learning process of pupils to apply cognitive and meta cognitive learning strategies in problem solving, memorizing and integrating new knowledge. Student teachers are for instance encouraged to use scaffolds in order to bridge the gap between what a pupil can do independently and the learning goals set. Other ways of bridging this gap might be modelling technics (thinking aloud) and control and checking strategies for the solutions found and to provide corrective feedback.

In Table 1 we provide an overview of the long standing research tradition into the categories and the references. Although research into the first three categories started decades ago, very little new research has been done recently, with the exception of some research into the ‘intellectual stimulating learning climate’ and ‘relations between pupils’. The results of this research therefore seem to be more or less finalized, because after all, we know what works and what does not. As shown, the last categories however are still objects of research.

For optimal support of student teachers through the different phases in their professional development, we updated the formal model didactical analysis (Van Gelder et al., 1971), based on teaching skills derived from these six categories of effective teacher behavior.

Part of the coaching approach we are developing is the use of a standardized observation instrument, in which the knowledge base on effective teaching behavior is operationalized. This instrument also has to be psychometrically tested on the existence of a hierarchical order in the teaching behavior measured.

The International Comparative Analysis of Learning and Teaching observation instrument (ICALT) meets these conditions. The ICALT observation instrument consists of six scales of effective teaching skills that are known to achieve involvement and learning gain of pupils (Van de Grift, Van der Wal, & Torenbeek, 1998; Van de Grift, 2014; Van de Grift & Lam, 1998). These scales are: (1) Creating a safe and stimulating learning climate, (2) Efficient classroom management, (3) Giving a clear instruction, (4) Activating pupils, (5) Adaptive teaching, and (6) Teaching learning strategies.

In several studies, using item response theory, the hierarchical order of complexity in teaching skills as measured with ICALT was established (Van Gelder et al., 1971). Therefore, the instrument can also help us in determining which teaching skills we should incorporate in each of the successive lesson templates.

2.3. Effective feedback

As mentioned in the introduction, the lesson templates currently used in the Netherlands serve as the main guidance instrument during internship. Unfortunately the templates lack sufficient guidelines for giving adequate feedback. For the development of the new lesson templates, we therefore utilize theory on effective feedback.

Hattie and Timperley (2007) find in 12 meta-analyses (with 196 studies) that feedback turns out to be one of the most important factors in achieving learning gain (average ES 0.79). The effect sizes measured however vary widely, which means that some forms of feedback are more powerful than others. Feedback appears to be more effective on a task level, when students receive feedback on the task cues (ES 1.10), and when they get suggestions on how to improve their task performance (0.94), compared to, for example, praise (ES 0.14), reward (ES 0.31) or punishment (ES 0.20). Hattie and Timperley developed a feedback model in order to improve learning gain. Their model aims to reduce the discrepancies between current understandings/performance and a desired goal (2007, p.7). Effective feedback therefore should accordingly answer three questions: ‘Where am I going?’ referring to the objectives (Feed up); ‘How am I going?’ referring to the progress made (Feedback) and ‘How next?’; referring to activities which will lead to better progress (Feedforward).

The focus on setting clear goals is also emphasized by Marzano (2007), who states that clear goals and giving feedback used in tandem, probably is more powerful than either of them applied in isolation, and that without clear objectives it might be difficult to provide students with effective feedback at all (p.12). In the current guidance situation however, student teachers have to formulate good and clear learning goals themselves. Unfortunately they are rarely successful in entering this first preparation step on the template. Mentors, due to the lack of feedback guidelines, mainly use subjective ad hoc feedback. Together they get caught up in what in literature is called a ‘feedback gap’, as mentor and student both are dissatisfied with the structure and outcome of the feedback cycle (Evans, 2013).

Incorporating Hattie and Timperley’s feedback model in our templates and also restructuring and standardizing this initial ‘goal setting’ preparation phase in a different manner on our restyled templates, should help to bridge this gap between the actual performance level and the desired learning objective (Lizzio & Wilson, 2008).

To support mentors further in guiding student teachers in this feedback cycle, we have to focus on the use of classroom observation and data driven feedback. This form of effective feedback links the theoretical knowledge offered at the teacher training college to the training practice at the internship schools. On the basis of classroom observations with a standardised observation instrument, student teachers can thus process their observation results actively in motivated and clearly indicated learning goals, whereas mentors can provide student teachers with effective feedback and teachers at the training college can more closely monitor their professional progress (Marsh, 2012; Van den Hurk, Houtveen, & Van de Grift, 2016).

In the next section we describe the further development of the lesson templates.

3. Pre-study

3.1. Development of the new lesson preparation and feedback templates

We developed four lesson preparation and feedback templates increasing in complexity; template 1 for Fuller and Bown’s ‘self-stage’, templates 2 and 3 for the ‘task-stage’ and template 4 for the ‘impact-stage’. Each template is divided into three components, namely design (A), actual lesson description (B) and feedback (C), in which the learning cycle of the pupils and the student teachers are combined.

In the design component the student explains the urgency to teach that lesson for that specific class at that moment by filling out elements

Table 1
Categories of effective teaching skills and references.

Categories	References
Safe and stimulating learning climate	
Safety (rules & regulations)	Evertson, Anderson, Anderson, and Brophy (1980); Gottfredson, Gottfredson, Payne, and Gottfredson (2005).
Learning climate	Brookover and Schneider (1975); Carver and Scheier (2000); Cohen, McCabe, Michelli, and Pickeral (2009); Deci and Ryan (2000); Evertson and Weinstein (2006); Evertson and Weinstein (2013); Rutter, Maughan, Moltre, Ouston, and Smith (1979); Wellisch, MacQueen, Carriere, and Duck (1978); Thapa, Cohen, Guffey, and Higgins-D'Alessandro (2013).
Relation (with teacher)	Davis (2003); Wigfield, Eccles, and Rodriguez (1998); Goodenow (1992); Osterman (2000); Perry and Weinstein (1998).
Relation (between pupils)	Finnan, Schnepel, and Anderson (2003); Ghaith (2003); Kerr, Ireland, Lopes, Craig, and Cleaver (2004).
Self-esteem & self-efficacy	Bandura (1986); Huebner and McCullough (2000); Stoel, Peetsma, and Roeleveld (2003).
High expectations	Dusek and Joseph (1983); Fuchs, Fuchs, Mathes, Lipsey, and Roberts (2001); Harris and Rosenthal (1985); Rosenthal and Rubin (1978); Smith (1980).
Efficient classroom management	
Effective organization and time use	Creemers (1994); Harnischfeger and Wiley (1978); Karweit (1989); Kindsvatter, Willen and Ishler, 1988; Rosenshine and Berliner (1978); Scheerens and Bosker (1997); Wang, Reynolds, and Walberg (1993).
Time-on-task	Mastropieri, Sweda, and Scruggs (2000); Van der Meer, Appelhof, ten Doesschate, and Kramer (1986); Veenman, Voeten, and Lem (1987).
Active time-on-task	Denham and Lieberman (1980); Fischer et al. (1980); Karweit (1982); Rosenshine and Berliner (1978).
Minimalize time loss	Bloom (1976); Brophy (1998); Doyle (1986); Harnischfeger and Wiley (1978); Wyne and Stuck (1982); Yair (2000).
Giving clear instruction	
Combining new with old knowledge	Creemers (1994); Scheerens (1992).
Lesson objectives, structure & oversight	Kameenui and Carnine (1998); Lohman (1986); Nunes and Bryant (1996); Pearson and Fielding (1991); Pearson and Gallagher (1983); Rosenshine and Meister (1997); Smith, Baker, Hattie, and Bond (2008).
Active teaching	
Correlation involvement & achievement	Denham and Lieberman (1980); Evertson et al. (1980); Fischer et al. (1980); Fredricks, Blumenfeld, and Paris (2004).
Intensifying instruction, through: Giving clear guidance	Harn, Linan-Thompson, and Roberts (2008)
Avoiding excessive seatwork	Hampton and Reiser (2004); Lang and Kersting (2007).
Higher order questions	Craig, Sullins, Witherspoon, and Gholson (2006); Mantione and Smead (2003); Murphy, Wilkinson, Soter, Hennessey, and Alexander (2009).
Cooperative learning	Meeuwisse, Severiens, and Born (2010).
Adaptive teaching	
Pre & re-teaching	Bosker (2005); Houtveen and Van de Grift (2012); Kidron and Lindsay (2014).
Adapting teaching & ability groups	Walmsley and Allington (1995); Houtveen, Booij, De Jong, and Van de Grift (1999); Slavin (1996).
Boosting self-esteem & motivation	Bennett (1986); Houtveen and Van de Grift (2007); Houtveen and Van de Grift (2001); Slavin and Braddock (1993).
Teaching learning strategies	
Scaffolds and corrective feedback	Dixon, Carnine, Lee, Wallin, and Chard (1998); Brophy and Good (1986); Hattie and Clinton (2008); Houtveen and Van de Grift (2007); Rosenshine and Stevens (1986); Slavin (1996); Smith et al. (2008).
Think aloud	Block and Israel (2004); Duffy (2003); Israel and Massey (2005); Pressley et al. (2002).

of the former didactical analysis model, such as: 'lesson goal', 'educational needs of the pupils' and 'evaluation'. In this component aspects of 'classroom and time management' are also placed. As mentioned above, the design elements increase in complexity. For example, in template 1 the accent is on setting clear rules and regulations that fit the self-stage and in template 4 the focus of the lesson design is on adaptive teaching, which fits the impact stage.

Finally, in the design component, the students indicate what instruction skill or skills they specifically want to practice in the prepared lesson. To give direction to the learning process of the student teacher, a selection of the ICALT items functions as a set of learning goals for the student. To do so, we matched the three stages of the concern theory with the following corresponding ICALT scales:

- The 'self-stage' with the scale 'Creating a safe and stimulating learning climate';
- The 'task-stage' with the scales 'Efficient classroom management' and 'Giving clear instruction';
- The 'impact-stage' with the scales 'Activating pupils', 'Adaptive teaching' and 'Teaching learning strategies'.

In this way student teachers practice only those instructional skills that match their stage of development.

The component 'actual lesson description' (B) replaces the former lesson description of the didactical analysis model. This component increases in complexity from aspects of 'direct instruction' in template 1, to 'adaptive instruction' in the last template. The student teachers fill out the activities they plan to perform and the activities they plan for the pupils.

The component feedback (C) is an addition in our templates. Following the aspects of effective feedback as described in the theoretical framework, the mentor gives feedback on the learning cycle of the children as well as the student teacher. The mentor observes the lesson, scores the instructional skills which the student teachers had planned to practice and writes down feedback based on these skills. Then the mentor gives feedback on the extent to which the learning goals for the pupils have been achieved and what consequences this has for the following lessons. The mentor also points out which instruction skills can best be practiced in the next lesson(s).

3.2. Content validity

The content validity of the lesson preparation templates was tested with an expert consultation. A panel of 50 professors, teacher trainers, internship supervisors, and student teachers gave their opinion by answering questions on the combination of the concern stage with the corresponding ICALT items, and the build-up in complexity in the templates. The panel also advised on the increased complexity in the lesson preparation elements 'lesson goals', 'educational needs of the pupils', 'lesson evaluation', 'time and classroom management' 'clear instruction', 'adaptive teaching', as well as the component 'evaluation' for the mentor. Because of the workability, the panel advised the number of templates should be reduced from four to three. A selection of experts advised on the content of the lectures at the teacher training college and the procedure in the schools. After this consultation, we adjusted the lesson preparation templates and put them to a test at the schools. The student teachers and supervisors who tested the templates approved of the workability. Minor adjustments resulted in the

Table 2
Design quasi-experiment (n = 198).

	Pre-test		Posttest
Experimental group	O ₁	X	O ₂
Control group	O ₁		O ₂

definitive version.

4. Method

The purpose of this research is to answer the question, whether student teachers who learn to teach with a coaching approach, involving classroom observation, assigning and using appropriate lesson preparation templates, combined with stage focused feedback at the school and lectures at the teacher training college, achieve a higher level of pedagogical and didactical teaching skills compared to student teachers who do not receive this approach.

4.1. Design

To answer the research question on the effectiveness of the coaching approach, a quasi-experiment with an untreated control group design with pre-test and posttest (Shadish, Cook, & Campbell, 2002) is used to rule out other variables that could be of influence (Table 2).

We measured the pedagogical and didactical teaching skills of the student teachers, by observing one of the first lessons during their internship (on a language or mathematics topic) and one of the last lessons of the academic year with the ICALT observation instrument (Van de Grift, 2014, 2007). This instrument consists of six scales: 'Creating a safe and stimulating learning climate' (4 items); 'Efficient classroom management' (4 items); 'Giving clear instruction' (7 items); 'Activating pupils' (7 items); 'Adaptive teaching' (4 items) and 'Teaching learning strategies' (6 items). The items can be scored on a 4-point scale: predominantly weak (0); more weaknesses than strength (1); more strength than weaknesses (2); and predominantly strong (3).

The experimental group received the treatment consisting of four elements: an initial observation to determine the best matching lesson preparation template; the use of the appropriate lesson preparation templates during internship; the stage focused feedback in the learning schools; and the attendance of lectures at the teacher training college. The control group received none of these treatment elements and used their conventional lesson preparation template.

4.2. Sample

The sample of this quasi-experiment involved 198 students; 135 in the experimental group, and 63 in the control group. The experimental group consisted of students from a teacher training college with locations in Utrecht and Amersfoort and the control group of students from comparable teacher training colleges elsewhere in the Netherlands.

We compared the samples of the control and experimental group. There was no significant difference between both groups in the percentage, whether it be student teachers' prior education, college year, their teaching a combination group or teaching kindergarten. There were, however, significant differences between the control group and the experimental group in the percentage of male students and regular college students. Because of these sample differences, it therefore is necessary to control for the variables 'percentage of male students' and 'regular students' to rule out the possible effect they have on the posttest results.

4.3. Treatment

The treatment consisted of four coherent elements. The first element

was the observation of a language or mathematics lesson and the assignment of the appropriate lesson preparation template. The actual use of the new lesson preparation templates during the internship, was the second element. The third element of the treatment was the stage focused feedback in the learning school. Student teachers received feedback specifically aimed at the teaching skills that the student should focus on during that period. The last element was the attendance at five lectures at the teacher training college. In the first lecture at the opening of the academic year, college teachers informed the students about the quasi-experiment, the new lesson preparation templates and the procedures. In the following lectures they introduced the correct use of the lesson preparation templates and the effective teaching skills. Furthermore, the students made a development plan of 'how to go to a higher level in teaching' and they received feedback on completed lesson preparation templates.

4.4. The procedure during internship

At the start of the student teacher's internship, there is an initial observation of a language or mathematics lesson with the ICALT observation instrument. On the basis of the ICALT observation instrument we developed the ICALT observation form. This ICALT observation form divides the ICALT items into three parts, which correspond with the three concern stages and indicate the matching lesson preparation template. Thus it helps assign the lesson preparation template that corresponds with the part of the observation form showing one or more insufficient items.

During the internship the elementary teacher (mentor) observes the lessons and scores the ICALT items on the lesson preparation template. After each lesson the mentor fills out the 'feedback' section on the lesson preparation template. The mentor evaluates whether the selected ICALT items have been sufficiently accomplished and what the consequences are for the next lesson(s). The mentor also points out which ICALT items can best be practiced. When all the ICALT items on the lesson preparation template score a three on a regular basis, the mentor and the ICALT observer can decide together that the student teacher is ready to move on to the next template.

4.5. Training

All teachers at the teacher training college, the ICALT observers and the mentors in the schools received the intended training. At the training college teachers employed standardized power points for the five lectures and additional digital training sessions. These sessions enabled us to store the students' replies for data feedback. We trained the ICALT observers in the correct use of the 'observation form' and the consequent assignment of the appropriate lesson preparation template, and the mentors in giving effective feedback on the lesson preparation template, using the component 'Feedback'.

Additionally, we trained all observers involved in the correct use of the ICALT observation instrument. During this training, we used two video recordings: a language and a mathematics lesson by a student teacher respectively in the third and fourth year of their teacher training.

Regarding the first video, 74% agreement over all the scales was reached upon the question of whether a teaching skill was sufficient/insufficient. With a Cohen's κ , this percentage of agreement was corrected for the probability of the correct answer. Seventy-four per cent corresponds with 0.48 Cohen's κ and indicates 'moderate agreement'. With the second training video the percentage of agreement increased to 81%. This corresponds with 0.62 Cohen's κ and indicates 'substantial agreement' (Cohen, 1960).

4.6. Data analysis

To measure the differences in development of teaching skills, we

Table 3

Comparison of the control group (n = 63) and the experimental group (n = 135).

	Control group	Experimental Group	Sign.
	%	%	
Male students	30.2	17.0	0.035
Prior education: Senior general secondary education (havo) & university preparatory education (vwo) (versus senior secondary vocational education (mbo))	76.2	72.9	n.s.
Regular student (versus part-time student)	98.4	89.6	0.030
College year 1	28.6	30.4	n.s.
College year 2	44.4	37.0	n.s.
College year 3	27.0	32.6	n.s.
Combination group (versus regular age group)	46.0	57.8	n.s.
Kindergarten (versus elementary education)	30.2	38.5	n.s.

used the ICALT observation instrument. We tested the reliability of the ICALT scales again with the ICALT scores of the pre-test and the posttest (Table 3). With an Analysis of Covariance (ANCOVA), we compared the average scores of the control group and the experimental group on the posttest, controlled for the variables pre-test, percentage male students and regular students. For the calculation of the effect size we used Morris's (2008) process, which defines the effect size as the difference between groups in the mean pre-post change, divided by the pooled pre-test standard deviation. According to Cohen (1988), effect sizes up to 0.19 can be seen as small effects, effect sizes from 0.50 to 0.79 as medium effects and 0.80 and above as large effects.

5. Results

5.1. Reliability

Testing the ICALT scales with the ICALT scores of the pre-test and the posttest (Table 4), secured that the reliability was high on the scales 'Safe learning environment', 'Clear instruction', 'Activating pupils', 'Adaptive teaching' and 'Learning strategies'. The reliability of the scale 'Classroom management' was acceptable.

5.2. Implementation of the treatment

To be able to attribute the effects of the quasi-experiment to the treatment, we studied the implementation of the four parts of the treatment: the observation and assignment of the appropriate lesson preparation template; the use of the appropriate templates; the stage focused feedback in the learning schools; and attendance at the lectures at the teacher training college. 100% of the observations and assignments were carried out. The use of the new templates was part of the internship assessment. The students uploaded their templates in a portfolio. We received 99.3% of the portfolios in semester 1 and 91.1% in semester 2.

Table 4

Reliability (Cronbach's Alpha), of the ICALT-scales on the pre-test and posttest (n = 198).

ICALT scales	Cronbach's Alpha	
	Pre-test	Posttest
1 Safe climate (4 items)	0.80	0.82
2 Classroom management (4 items)	0.70	0.74
3 Clear instruction (7 items)	0.85	0.88
4 Activating pupils (7 items)	0.85	0.85
5 Adaptive teaching (4 items)	0.81	0.88
6 Learning strategies (6 items)	0.89	0.88

We also checked the portfolios for authenticity and for the correct amount of templates. Although the percentage of the portfolios in semester 2 decreased slightly, the amount of lesson preparation templates per portfolio increased. All portfolios in semester 2 met the pre-determined standard of, depending on the year of training, 20–30 templates per semester. A total of 86.7% of the students received feedback on the templates. The use of logbook registrations and attendance lists confirmed that the teachers at the college gave all due 'lectures' as intended and that 81.4% of the student teachers attended at least four out of five lectures. We can conclude that the treatment was implemented sufficiently.

6. Effects

In the comparison of the pre-tests, the average score of the scale 'Adaptive teaching' was significantly higher in the experimental group. The other five ICALT scales did not show a significant difference between the pre-test of the control group and the experimental group. This meant that besides controlling for the above-mentioned variables male students and regular college students, we also had to control for the pre-test.

Furthermore, we compared the difference of the average scores on the six ICALT-scales of the control group in the pre-test and the posttest situation. The differences found between the pre-test and posttest are small and not significant (not shown in Table 5). The average scores on the posttest of the experimental group are significantly higher on all six ICALT scales compared to the pre-test (not shown in Table 5).

In the previous sections we established that there are differences between the control group and the experimental group with regard to the percentage of male students, of regular college students and the scale 'Adaptive teaching' in the pre-tests. Therefore we used an ANCOVA procedure to control for these variables on all six ICALT scales. The differences found between the control group and the experimental group on the posttest, after controlling for pre-test, the percentage of male students and regular college students, were significant on all the six ICALT scales ($p < .001$). (See the penultimate column in Table 5).

6.1. The effect sizes

The effect size is measured by a procedure that divides the difference between both groups in the average pre-post change by the pooled pre-test standard deviation (Morris, 2008). In our case, these effect sizes were respectively: 0.74, 0.85, 0.69, 0.97, 1.16, and 0.96 (cf. the last column of Table 4). According to Cohen (1988), these effect sizes found were medium (0.50–0.79) for the scales 'Creating a safe learning climate' and 'Giving a clear instruction', and large (≥ 0.80) for the scales 'Efficient classroom management', 'Activating pupils', 'Adaptive teaching' and 'Teaching learning strategies'.

7. Conclusion and discussion

The purpose of this research was to answer the question, whether student teachers who learn to teach with a coaching approach, involving classroom observation, assigning and using appropriate lesson preparation templates, combined with stage focused feedback at the school and lectures at the teacher training college, achieve a higher level of pedagogical and didactical teaching skills compared to student teachers who do not receive this approach. We developed a coaching approach consisting of these four elements and conducted a quasi-experiment with an untreated control group design with a pre-test and posttest. After establishing that the treatment implementation was sufficient, we used an ANCOVA procedure to compare the differences between the control group and the experimental group on the posttest, after controlling for the pre-test, and the percentage of male and regular college students. The differences were significant on all the six ICALT

Table 5
ICALT results of the pre-test and posttest in the control group and experimental group (n = 198).

n	Control group				Experimental group				Difference	
	Pre-test		Posttest		Pre-test		Posttest			
	63		63		135		135			
ICALT	M	(SD)	M	(SD)	M	(SD)	M	(SD)	Sign	ES
Safe climate	2.27	(0.58)	2.27	(0.64)	2.23	(0.52)	2.63	(0.46)	0.000	0.74
Classroom management	1.73	(0.61)	1.80	(0.52)	1.71	(0.63)	2.31	(0.60)	0.000	0.85
Clear instruction	1.76	(0.60)	1.87	(0.59)	1.77	(0.61)	2.30	(0.55)	0.000	0.69
Activating pupils	1.27	(0.59)	1.33	(0.59)	1.28	(0.77)	2.04	(0.61)	0.000	0.97
Adaptive teaching	0.47	(0.58)	0.50	(0.64)	0.82	(0.79)	1.70	(0.85)	0.000	1.16
Learning strategies	0.74	(0.57)	0.88	(0.62)	0.86	(0.78)	1.69	(0.70)	0.000	0.96

scales. These effect sizes were medium for the scales ‘Creating a safe learning climate’ (.74), and ‘Giving clear instruction’ (.69), and large for the scales ‘Efficient classroom management’ (.85) ‘Activating pupils’ (.97), ‘Adaptive teaching’ (1.16) and ‘Teaching learning strategies’ (.96). On account of these results we can conclude that student teachers who learned to teach with this coaching approach achieved a higher level of pedagogical and didactical teaching skills.

It deserves notice that the control group shows no significant growth of the average ICALT scales between the pre-test and the posttest. In an attempt to understand why the students hardly advanced in pedagogical and didactical skills, we studied the internship assessment indicators. In the Netherlands the internship assessment indicators used are largely based on the standards of the Foundation for Professional Qualities of Teachers (Stichting Beroepskwaliteit Leraren), the so-called ‘SBL competences’ (2004). Despite of recent adjustments of these standards by the *Educational Cooperative* (2014) (Onderwijs coöperatie), the internship assessment lists used appear to contain few or no indicators for observing effective teaching skills in the daily classroom situation. Only three out of seven SBL competence categories deal with teaching matters in the classroom and these competences are formulated in a rather broad and general way. It is conceivable that these assessment lists offer assessors too little direction in evaluating effective teaching skills and student teachers too little feedback on developing these skills. The fact that the control group hardly advanced might be attributed to this lack of direction. Further research on these particular assessment indicators in relation to acquiring effective teaching skills would be interesting.

The collaboration between the teacher training college and the internship schools involved needs some consideration. Our coaching approach consisted of the application of a procedure in which the teacher training college and internship schools collaborate in collecting objective data on student teaching skills and in presenting coherent guidelines for giving effective feedback for both teacher training college and internship schools alike. Applying this coaching approach demanded a close collaboration of participants in both work fields, as the teacher training college had to readdress the content of its curriculum, and the internship schools as well. Strikingly however, the internship schools and their management boards took the lead in the innovation, during the implementation and research period. This readiness to join the new approach finds a possible explanation in sheer necessity, because the process of learning to teach predominantly takes place in practice and in it the lack of guidelines has become rather pressing. The internship schools thus readily embraced the lesson templates and participated fully in the ICALT training. While the teacher training college, already struggling with an overburdened curriculum, was reconsidering its involvement, it was their enthusiasm that inspired the teacher training college to collaborate further, by pointing out the importance of the innovation at hand. The strength of this collaboration in combination with the results found, has led to the joint decision of the management of the teacher training college and the internship schools to implement

the new coaching approach in the curriculum.

To understand more fully which parts of the treatment especially attributed to the effects found on the posttest of the six ICALT scales, we performed several multilevel regression analyses. We analyzed the impact quality of the four parts of the treatment and the pre-test in six models, on all six ICALT scales. In every model we included a separate element of the treatment (‘observation and assignment’, ‘use of the lesson preparation templates’, ‘attending lectures’ and ‘mentor feedback’) and in the last model we added all elements. The models that analyzed the elements separately offered a significant explanation with regression coefficients (β) from .38 to .81. However, when we included all elements of the treatment in model six, the first element we added explained all the variance and none of the other treatment variables sorted any significant effect. This can be explained by the very high correlations between the individual treatment elements. The inter correlations are all over the .90, which is an indication of the very strong internal consistency of the treatment; after all it is quite impossible to do one thing without the other. Because of this high correlation, it is impossible to analyze if ‘mentor feedback’ is especially responsible for the effects found, because without observations and assignments, there is no occasion for conducting other elements of the treatment. To understand the individual impact of the elements, it would be interesting to conduct an experiment with a treatment which leaves out one or more elements.

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.stueduc.2018.05.005>.

References

- 10voordeleraar (2018). *Kennisbases en profilering lerarenopleiding basisonderwijs*. [Knowledge base and profiling teacher training college for elementary education]. Retrieved from: <https://www.10voordeleraar.nl/publicaties>.
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, 4(3), 359.
- Bennett, C. I. (1986). *Comprehensive multicultural education: Theory and practice*. Boston: Allyn and Bacon.
- Block, C. C., & Israel, S. E. (2004). The ABCs of performing highly effective think-alouds. *The Reading Teacher*, 58(2), 154–167.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Bosker, R. J. (2005). *De grenzen van gedifferentieerd onderwijs*. [The limits of adaptive education]. Groningen: Inaugurele rede. Rijksuniversiteit Groningen.
- Boz, Y. (2008). Turkish student teachers' concerns about teaching. *European Journal of Teacher Education*, 31(4), 367–377.
- Broad, K., & Evans, M. (2006). *A review of literature on professional development content and delivery modes for experienced teachers*. University of Toronto, Ontario Institute for Studies in Education.
- Brookover, W. B., & Schneider, J. M. (1975). Academic environments and elementary

- school achievement. *Journal of Research & Development in Education*.
- Brophy, J. (1998). Classroom management as socializing students into clearly articulated roles. *The Journal of Classroom Interaction*, 1–4.
- Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.). *Handbook of research on teaching* (pp. 328–375). New York: Macmillan.
- Buchmann, M. (1987). Role over person, morality and authenticity in teaching: Justifying teacher action and decisions. *Scandinavian Journal of Educational Research*, 31(1), 1–21.
- Carver, C. S., & Scheier, M. F. (2000). *On the structure of behavioral self-regulation. Handbook of self-regulation* 41–84.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *The Teachers College Record*, 111(1), 180–213.
- Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, 24(4), 565–591.
- Creemers, B. P. (1994). *The effective classroom*. London: Cassell.
- Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist*, 38(4), 207–234.
- Day, C., & Gu, Q. (2010). *The new lives of teachers*. London: Routledge.
- Day, C., & Qing, G. (2007). Variations in the conditions for teachers' professional learning and development: Sustaining commitment and effectiveness over a career. *Oxford Review of Education*, 33(4), 423–443 Taylor and Francis group.
- Deci, E. L., & Ryan, R. M. (2000). The support of autonomy and the control of behavior. In E. T. Higgins, & A. W. Kruglanski (Eds.). *Key reading in social psychology. Motivational science: Social and personality perspectives* (pp. 128–145). New York, NY, US: Psychology Press.
- Denham, C., & Lieberman, A. (1980). *Time to learn. A review of the beginning teacher evaluation study*.
- Dixon, R., Carnine, D., Lee, D., Wallin, J., & Chard, D. (1998). *Review of high-quality experimental mathematics research: Report to the California State board of education* Eugene, OR: National Center to Improve the Tools of Educators.
- Doyle, W. (1986). Classroom organization and management. *Handbook of Research on Teaching*, 3, 392–431.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Oxford: Basil Blackwell.
- Duffy, G. G. (2003). *Explaining reading: A resource for teaching concepts, skills, and strategies*. Newark, DE: International Reading Association.
- Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, 75(3), 327.
- Ellis, E. S., & Worthington, L. A. (1994). *Research synthesis on effective teaching principles and the design of quality tools for educators. Technical report no. 5*.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120.
- Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American Educational Research Journal*, 17(1), 43–60.
- Evertson, C. M., & Weinstein, C. S. (2006). *Classroom management as a field of inquiry. Handbook of classroom management: Research, practice, and contemporary issues*, 3, 16.
- Evertson, C. M., & Weinstein, C. S. (2013). *Handbook of classroom management: Research, practice, and contemporary issues*. Mahwah, NJ: Laurence Erlbaum Associates.
- Fessler, R., & Rice, E. (2010). Teachers career stages and professional development. In P. Peterson, E. Baker, & B. McGaw (Vol. Eds.), *International encyclopedia of education: Vol. 3*, (pp. 582–586). Oxford: Elsevier.
- Finnan, C., Schnepel, K. C., & Anderson, L. W. (2003). Powerful learning environments: The critical link between school and classroom cultures. *Journal of Education for Students Placed at Risk*, 8(4), 391–418.
- Fischer, C., Berliner, D., Filby, N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham, & A. Lieberman (Eds.). *Time to learn* Washington DC: Department of Education and the National Institute of Education pp. 7e32.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Lipsey, M. W., & Roberts, P. H. (2001). Is "learning disabilities" just a fancy term for low achievement? A meta-analysis of reading differences between low achievers with and without the label. *Paper Written for the Office of Special Education Programs*.
- Fuller, F. F., & Bown, O. H. (1975). Becoming a teacher. In K. Ryan (Ed.). *Teacher education (74th yearbook of the national society for the study of education) part 2* (pp. 25–52). Chicago: University of Chicago Press.
- Ghaith, G. (2003). Effects of the learning together model of cooperative learning on English as a foreign language reading achievement, academic self-esteem, and feelings of school alienation. *Bilingual Research Journal*, 27(3), 451–474.
- Ghaith, G., & Shaaban, K. (1999). The relationship between perceptions of teaching concerns, teacher efficacy, and selected teacher characteristics. *Teaching and Teacher Education*, 15(5), 487–496.
- Goodenow, C. (1992). *School motivation, engagement, and sense of belonging among urban adolescent students*.
- Goncalves, J. A. (2009). Professional development and teaching career. Career phases, curriculum and supervision. *Sisifo Educational Science Journal*, 8, 21–32.
- Gottfredson, G. D., Gottfredson, D. C., Payne, A. A., & Gottfredson, N. C. (2005). School climate predictors of school disorder: Results from a national study of delinquency prevention in schools. *Journal of Research in Crime and Delinquency*, 42(4), 412–444.
- Hampton, S. E., & Reiser, R. A. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Research in Higher Education*, 45(5), 497–527.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466–479.
- Harn, B. A., Linan-Thompson, S., & Roberts, G. (2008). Intensifying instruction: Does additional instructional time make a difference for the most at-risk first graders? *Journal of Learning Disabilities*, 41(2), 115–125.
- Harnischfeger, A., & Wiley, D. E. (1978). Conceptual issues in models of school learning. *Journal of Curriculum Studies*, 10(3), 215–231.
- Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97(3), 363.
- Hattie, J. A. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.
- Hattie, J., & Clinton, J. (2008). Identifying accomplished teachers: A validation study. In L. C. Ingvarson, & J. Hattie (Vol. Eds.), *Assessing teachers for professional certification: The first decade of the national board for professional teaching standards: Vol. 11*, (pp. 313e–). Amsterdam: Elsevier Press.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Houtveen, A. A. M., & Van de Grift, W. J. C. M. (2001). Inclusion and adaptive instruction in elementary education. *Journal of Education for Students Placed at Risk*, 6(4), 389–409.
- Houtveen, A. A. M., Booi, N., De Jong, R., & Van de Grift, W. J. C. M. (1999). Adaptive instruction and pupil achievement. *School Effectiveness and School Improvement*, 10(2), 172–192.
- Houtveen, A., & Van de Grift, W. (2007). Effects of metacognitive strategy instruction and instruction time on reading comprehension. *School Effectiveness and School Improvement*, 18(2), 173–190.
- Houtveen, A. A. M., & Van de Grift, W. J. C. M. (2012). *Differentiatie in het basisonderwijs. Meer leerwinst door betere afstemming. [More learning gain with better adaptive teaching skills]. Schoolmanagement totaal, september 2012*. 26–29.
- Huberman, M. (1989). On teachers' careers: Once over lightly, with anbrood brush. *International Journal of Educational Research*, 13(4), 347–362.
- Huebner, E. S., & McCullough, G. (2000). Correlates of school satisfaction among adolescents. *The Journal of Educational Research*, 93(5), 331–335.
- Israel, S. E., & Massey, D. (2005). *Metacognitive think-alouds: Using a gradual release model with middle school students. Metacognition in literacy learning: Theory, assessment, instruction, and professional development* 183–198.
- Kameenui, E. J., & Carnine, D. (Eds.). (1998). *Effective teaching strategies that accommodate diverse learners*. Upper Saddle River, NJ: Prentice Hall.
- Karweit, N. (1982). *Time on task: A research review. Report no. 332*.
- Karweit, N. (1989). The effects of a story-reading program on the vocabulary and story comprehension skills of disadvantaged prekindergarten and kindergarten students. *Early Education and Development*, 1(2), 105–114.
- Kerr, D., Ireland, E., Lopes, J., Craig, R., & Cleaver, E. (2004). *Making Citizenship education real: Citizenship education longitudinal study. Second annual report* London: DfES.
- Kidron, Y., & Lindsay, J. (2014). *The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta-analytic review (REL 2014-015)*. Washington DC: Regional Educational Laboratory Appalachia.
- Kindsavatter, Willen, R. W., & Ishler, M. (1988). *Dynamics of effective teaching*, 20–34, New York: Longman 20–34.
- Lang, J. W., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, 35(3), 187–205 (Eds.), *Handbook of research on multicultural education* (pp. 525–547). New York: Macmillan.
- Levine, D. U., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*.
- Levine, D. U., & Lezotte, L. W. (1995). Effective schools research. In J. A. Banks, & C. A. M. Banks (Eds.). *Handbook of research on multicultural education* (pp. 525–547). New York: Macmillan.
- Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263–275.
- Lohman, D. L. (1986). Predicting mathemathanic effects in the teaching of higher-order thinking skills. *Educational Psychologist*, 21(3), 191–208.
- Mantione, R. D., & Smead, S. (2003). *Weaving through words: Using the arts to teach reading comprehension strategies*. Newark, DE: International Reading Association.
- Marsh, J. (2012). Interventions promoting educators' use of data: Research insights & gaps. *Teachers College Record*, 114(11), 1–48.
- Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: ASCD.
- Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: ASCD.
- Mastropieri, M. A., Sweda, J., & Scruggs, T. E. (2000). Putting mnemonic strategies to work in an inclusive classroom. *Learning Disabilities Research & Practice*, 15(2), 69–74.
- Meelissen, M., & Punter, A. (2016). *Twintig jaar TIMSS. Ontwikkelingen in leerling prestaties in de Exacte vakken in het basisonderwijs 1995-2015. [Twenty years of TIMSS: Developments in student performance in the exact subjects in elementary education]*. Enschede: IEA TIMSS & PIRLS, Universiteit Twente.
- Meeuwisse, M., Severiens, S. E., & Born, M. P. (2010). Learning environment, interaction, sense of belonging and study success in ethnically diverse student groups. *Research in Higher Education*, 51(6), 528–545.
- Meijerink, H. (2012). *Een goede basis. [A good base]. Commissie kennisbases*.
- Ministerie van onderwijs cultuur en wetenschappen (2013). *Lerarenagenda 2013-2020: de*

- leraar maakt het verschil. [Teacher agenda: The teacher makes the difference]. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap (OCW).
- Morris, S. B. (2008). Estimating effect sizes from pre-test–posttest–control group designs. *Organizational Research Methods*, 11(2), 364–386.
- Murphy, P. K., Wilkinson, I. A., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, 101(3), 740.
- Muijs, D., & Reynolds, D. (2010). *Effective teaching: Evidence and practice*. London: Sage.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Wiley-Blackwell.
- OECD (2011). *Education at a glance 2011: OECD indicators*. OECD Publishing <http://dx.doi.org/10.1787/eag-2011-en>.
- Onderwijs coöperatie. (2014, juni). *Herijking bekwaamheidseisen*. [Calibrating the competence standards]. <https://www.onderwijscooperatie.nl/wp-content/uploads/sites/2/2017/10/Bekwaamheidseisen-ingangsdatum-1-augustus-2017.pdf>.
- Osterman, K. F. (2000). Students' need for belonging in the school community. *Review of Educational Research*, 70(3), 323–367.
- Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Vol. Eds.), *Handbook of reading research: Vol. 2*, (pp. 815–860). New York: Longman.
- Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology*, 8(3), 317–344.
- Perry, K. E., & Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist*, 33(4), 177–194.
- Pigge, F. L., & Marso, N. R. (1997). A seven-year longitudinal multi factor assessment of teaching concern preparation and early years of teaching. *Teaching and Teacher Education*, 13(2), 225–235.
- Pressley, M. (2002). Comprehension strategies instruction. In C. Block, & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 11–27). New York: Guilford.
- Reeves, C. K., & Kzelskis, R. (1985). Concerns of preservice and inservice teachers. *Journal of Educational Research*, 78(5), 267–271.
- Richardson, V., & Placier, A. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research edition*. Washington DC: American Educational Research Association.
- Rosenshine, B. V., & Berliner, D. C. (1978). Academic engaged time. *British Journal of Teacher Education*, 4(1), 3–16.
- Rosenshine, B. V., & Meister, C. (1997). Cognitive strategy instruction in reading. In S. A. Stahl, & D. A. Hayes (Eds.), *Instructional models in reading* (Hillsdale, NJ: Lawrence Erlbaum pp. 85e).
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 1(3), 377–386.
- Rosenshine, B., & Stevens (1986). Teaching functions. *Handbook of Research on Teaching*, 376–391.
- Rutter, M., Maughan, B., Moltre, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their affects on children*. Cambridge: Harvard University Press.
- Scheerens, J. (1992). *Effective schooling: Research theory and practice*. London: Cassell.
- Scheerens, J. (2015). *School effectiveness research*. In *International encyclopedia of social and behavioral sciences* (2nd edition). Elsevier.
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Pergamon.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Slavin, R. E. (1996). *Education for all*. Lisse, the Netherlands: Swets & Zeitlinger.
- Slavin, R. E., & Braddock, J. H. I., II (1993). Ability grouping: On the wrong track. *College Board Review*, 168, 11–17.
- Smith, M. L. (1980). Teacher expectations. *Evaluation in Education*, 4, 53–55.
- Smith, D. J., & Sanche, R. P. (1992). Saskatchewan interns' concerns at three stages of a four-month practicum. *Alberta Journal of Educational Research*, 38(2), 121–132.
- Smith, T., Baker, W., Hattie, J., & Bond, L. (2008). A validity study of the certification system of the national board for professional teaching standards. In L. C. Ingvarson, & J. Hattie (Vol. Eds.), *Assessing teachers for professional certification: The first decade of the national board for professional teaching standards: Vol. 11*, (pp. 345–380). Amsterdam: Elsevier Press.
- Steffy, B. E., & Wolfe, M. P. (2001). A life cycle model for career teachers. *Kappa Delta Pi Records*, 38, 16–19.
- Stichting Beroepskwaliteit Leraren en ander onderwijspersoneel (SBL) [Foundation for Professional Qualities of Teachers] (2004). *Bekwaamheidseisen leraren. Eindversie 20 mei 2004*. Den Haag: SBL.
- Stoel, R. D., Peetsma, T. T., & Roeleveld, J. (2003). Relations between the development of school investment, self-confidence and language achievement in elementary education: A multivariate latent growth curve approach. *Learning and Individual Differences*, 13(4), 313–333.
- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83(3), 357–385.
- Van de Grift, W. J. C. M. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152.
- Van de Grift, W. J. C. M. (2014). Measuring teaching quality in several European countries. *School Effectiveness and School Improvement*, 25(3), 295–311.
- Van de Grift, W. J. C. M., & Lam, J. (1998). Het didactisch handelen in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 23(2), 224–241.
- Van de Grift, W. J., Van der Wal, M., & Torenbeek, M. (2011). Ontwikkeling in de pedagogisch didactische vaardigheden in het basisonderwijs. *Pedagogische studiën*, 88, 416–432.
- Van Gelder, L., Oudkerk Pool, T., Peters, J., & Sixma, J. (1971). *Didactische analyse*. Groningen: Wolters-Noordhoff.
- Van den Hurk, W. H. T., Houtveen, A. A. M., & Van de Grift, W. J. C. M. (2016). Fostering effective teaching behavior through the use of data-feedback. *Teaching and Teacher Education*, 60, 444–451.
- Van der Meer, T. A. M., Appelhof, P. N., ten Doesschate, J. J., & Kramer, G. B. (1986). *Het onderwijs in de klas tussen 9 en 12 [Education for children between 9 and 12 years of age]* (Utrecht, school Advies Centrum/Vakgroep Onderwijskunde).
- Veenman, S., Voeten, M., & Lem, P. (1987). Classroom time and achievement in mixed age classes. *Educational Studies*, 13(1), 75–89.
- Walmsley, S. A., & Allington, R. L. (1995). Redefining and reforming instructional support programs for at-risk students. In R. L. Allington, & S. A. Walmsley (Eds.), *No quick fix: Rethinking literacy programs in America's elementary schools* (pp. 19–44). Newark, DE: International Reading Association.
- Wang, M. C., Reynolds, M. C., & Walberg, H. J. (1993). What works and what doesn't work: The case for an inclusive system. *National Center on Education in the Inner Cities*, 32 #93-8. ED 399 316. UD 031 278.
- Watzke, J. L. (2003). Longitudinal study of stages of beginning teacher development in a field-based teacher education program. *The Teacher Educator*, 38(3), 209–229.
- Watzke, J. L. (2007). Longitudinal research on beginning teacher development: Complexity as a challenge to concerns-based stage theory. *Teaching and Teacher Education*, 23(1), 106–122.
- Wellisch, J. B., MacQueen, A. H., Carriere, R. A., & Duck, G. A. (1978). School management and organization in successful schools (ESAA in-depth study schools). *Sociology of Education*, 211–226.
- Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. Chapter 3 *Review of Research in Education*, 23(1), 73–118.
- Woods, A. M., & Lynn, S. A. (2001). Through the years: A longitudinal study of physical education teachers from a research-based preparation program. *Research Quarterly for Exercise and Sport*, 72(3), 219–231.
- Wyne, M. D., & Stuck, G. B. (1982). Time and learning: Implications for the classroom teacher. *The Elementary School Journal*, 83(1), 67–75.
- Yair, G. (2000). Educational battlefields in America: The tug-of-war over students' engagement with instruction. *Sociology of Education*, 247–269.
- Zeichner and Teitelbaum (1982). Alternative paradigms to teacher education. *Journal of Teacher Education*, 34(3), 32–48.

Tamar Tas is a researcher and a teacher educator at the Department of Education of the Utrecht University of Applied Sciences in the Netherlands. Her key area of interest is improving the quality of teaching skills of student teachers, curriculum development, and improving the quality of guidance in the internship schools.

Thoni Houtveen is a professor of literacy at the Department of Education of the Utrecht University of Applied Sciences in the Netherlands. Her key areas of interest are the effects of innovations in primary and secondary education and the evaluation of school improvement programs aimed at the improvement of language/reading education in primary and special education.

Wim van de Grift is an emeritus professor of Education and director of the Department of Teacher Education of the Groningen University in the Netherlands. Furthermore he was a scientific advisor at inspectorate of education in the Netherlands and member of the Dutch Program Council for Educational Research of the Netherlands Organization for Scientific Research.

Melissa Willemsen is a researcher at the Department of Education of the Utrecht University of Applied Sciences in the Netherlands.