

De invloed van docentenprofessionalisering op het leren van studenten

Paper te presenteren op Studiedag VFO, UHasselt, 13 november 2008

Ann Stes^o

Sven De Maeyer

David Gijbels

Peter Van Petegem

Universiteit Antwerpen

^o corresponderende auteur:

Ann Stes, Universiteit Antwerpen, Instituut voor Onderwijs- en Informatiewetenschappen,
ExpertiseCentrum Hoger Onderwijs (ECHO), Venusstraat 35, BE-2000 Antwerpen.

E-mail: ann.stes@ua.ac.be

Niet citeren zonder toestemming van de auteurs

Abstract

Onderzoeksvragen

- 1) Is er een invloed van docentenprofessionalisering op het leergedrag van studenten?
- 2) Is er een invloed van docentenprofessionalisering op de leerresultaten van studenten?
- 3) Hangt de invloed van docentenprofessionalisering op het leergedrag en de leerresultaten van studenten af van de groepsgrootte en het niveau (eerstejaars versus niet eerstejaars) van de studenten?

Onderzoeksopzet

Voor de start (pre-test) zowel als na afloop (post-test) van een onderwijskundig professionaliseringstraject voor beginnende docenten hoger onderwijs werden er kwantitatieve data verzameld bij telkens ruim 1000 studenten en dit volgens een quasi-experimenteel design (met 19 docenten behorend tot de experimentele groep, 18 tot de controlegroep). Voor de bevraging van de studenten werd gebruik gemaakt van voor het onderzoek gevalideerde vragenlijsten (R-SPQ-2F en deel 4 van ETLQ). Een multi-level analyse werd uitgevoerd op de data waarbij vijf modellen werden geschat.

Belangrijkste resultaten

Een eerste basismodel gaf aan dat docenten van elkaar verschillen wat betreft het leergedrag en de leerresultaten van hun studenten. Anderzijds vloeien verschillen in schaa scores ook in belangrijke mate voort uit verschillen tussen studenten. Een tweede model - waarin de mogelijke interveniërende invloed van docent-, context- en studentkenmerken buiten beschouwing werd gelaten - duidde geen significant effect aan van docentenprofessionalisering op het leergedrag en de leerresultaten van studenten. Een derde model bracht een zekere netto impact van professionalisering aan het licht; deze was, opmerkelijk, negatief. Een eerste interactiemodel bevestigde dat de impact van docentenprofessionalisering op het leergedrag en de leerresultaten van studenten globaal positiever is voor docenten die lesgeven aan eerstejaars dan voor hen die lesgeven aan niet eerstejaars. Een laatste model gaf aan dat de impact van docentenprofessionalisering ook afhankelijk is van de groepsgrootte: globaal is de impact op het leergedrag en de leerresultaten positiever naarmate de groep studenten groter is.

Introduction

In recent years teacher training in higher education has become an important concern. However, despite the importance attached to this topic, evaluations have generally been limited to measuring participants' satisfaction and little is known about the impact on day-to-day teaching practice (Eison & Stevens, 1995; Norton, Richardson, Hartley, Newstead & Mayes, 2005; Wilson & Berne, 1999). What teachers actually learn from instructional development also remains unclear (Fishman, Marx, Best & Tal, 2003). Gibbs and Coffey (2000) conclude that there is still very little evidence that training university teachers makes any difference. The conceptual underpinning of what little empirical evidence there is also leaves a good deal to be desired and the lack of systematic program evaluation is an on-going concern in instructional development (Gibbs & Coffey, 2004; Gilbert & Gibbs, 1999; Ho, 1998; Postareff, Lindblom-Ylänne & Nevgi, 2007; Rust, 1998).

Existing studies on the impact of teacher training in higher education have primarily focused on the impact on teaching conceptions, beliefs or intentions (e.g. Ho, Watkins and Kelly, 2001; Norton et al., 2005; Stes, Clement and Van Petegem, 2007a). Research has also been conducted into the impact of training on teaching approach (e.g. Gibbs and Coffey, 2004; Postareff, Lindblom-Ylänne & Nevgi, 2007; Stes, Coertjens and Van Petegem, 2007b). Most studies indicate that teacher training has a positive effect.

Impact of teacher training in higher education on teaching behavior

Besides measuring the effects of training on teaching conceptions, intentions, beliefs and approaches, it is also worth considering effects on teaching behavior. Indeed, the relationship between teachers' conceptions, beliefs and approaches to teaching, on the one hand, and their actual teaching behavior when giving particular courses, on the other hand, is not as straight-forward as it may seem (Murray and McDonald, 1997; Trigwell and Prosser, 1996). Given that teaching takes place within a particular context (defined by elements such as the number and ability of the students, support from colleagues, the teacher's work load and teaching experience, etc.), the freedom of an instructor to teach in accordance with how he or she conceives teaching, is undoubtedly limited. Moreover, not only the teaching context itself, but also the teacher's perception of that context may affect the way he or she acts (Prosser & Trigwell, 1997).

In their research into the effectiveness of teacher training Stes et al. (2007a) found a long term impact in terms of individual teaching behavior. Brawner, Felder, Allen and Brent (2002) came to the conclusion that teachers participating in training programs were more convinced of the importance of active learning and behaved differently as a consequence: for example, by giving more class-based group assignments. An earlier study by Rust (1998) revealed that instructional development workshops promote at least some changes in the teaching practice of most participants and extensive change in some participants. All changes were regarded as successful.

Involving students in impact research

In their study of the impact of teacher training on teacher behavior Ho et al. (2001) also involved students. This is certainly a worthwhile approach given that the way students perceive teaching affects their learning (Entwistle, 1998). Students' perceptions of the learning environment have an impact on the way they study and thus influence learning results. Furthermore, the educational environment itself is, to a large extent, created by students' experience of curricula, teaching methods and assessment procedures. Students respond to the situation they perceive, which is not necessarily the same as the situation that their teachers have defined. This also explains why the effects of a new teaching and learning context on student learning are often the opposite of those intended by its designers, precisely because the students concerned see things differently (Ramsden, 1997). Marked contrasts were found between instructor and student reporting of teaching strategies used in lectures (Willcoxson, 1998). Student feedback is helpful because students are actively engaged in the teaching and learning process and can, therefore, provide an accurate evaluation of the teacher's classroom behavior (Ramsden, 1991). Ho et al. (2001) found that all teachers who changed their conceptions of teaching as a result of having followed a staff development program also received better ratings on their teaching practice from their students in the following academic year. Likewise, Coffey and Gibbs (2000) and Gibbs and Coffey (2004) also involved students in their research into the impact of teacher training and concluded that instructors significantly benefited from training when scores on a student rating questionnaire were considered. Stes et al. (2007b) gathered quantitative and qualitative student data taken from both an experimental and a control group in order to examine the impact of a staff development program on instructors' teaching behavior as perceived by students. The quantitative data were aggregated to teacher level, thus internal differences between students were not taken into account. No significant differences between the experimental and control groups were revealed. Similarly, the qualitative data obtained from semi-structured interviews did not point to any effect of training on teaching behavior.

Impact of teacher training in higher education on student learning

In spite of these pivotal studies, it is nonetheless apparent that the evidence regarding the impact of teacher training in higher education is so far still limited to data relating to instructor's teaching. Evidence with respect to the effects on student learning is almost non-existent (Ho et al., 2001). As Gibbs and Coffey (2004, p.88) point out:

“Studies tend not to obtain evidence from theoretically or psychometrically based questionnaires, obtain evidence from students or obtain evidence about the impact on student learning, which is the ultimate purpose of training”.

There is a clear need for research into the impact of instructional development on student learning and achievement (Penuel, Fishman, Yamaguchi & Gallagher, 2007; Stes et al., 2007b).

The research carried out by Gibbs and Coffey (2004) was the first published study to make use of psychometric student data and to include a control group so as to be able to measure the impact of teacher training in higher education. A questionnaire was put to students at the start of their teacher's training and again one year later in 22 universities in eight countries. A control group of students at two other universities which offered no

training or organized support for teachers was studied in the same way. An analysis of the quantitative data revealed that training not only improves students' perceptions of teaching but also their approach to learning.

We are not aware of any previous research into the impact of training on students' learning outcomes. However, a study by Lea, Stephenson and Troy (2003) reveals that learning environments in higher education that are based on student-centered learning have the potential to improve student learning outcomes. It would, thus, be interesting to look at whether training aimed at making the participants' teaching practice better geared to the concept of student-centered learning has a positive impact on student learning outcomes.

Factors influencing the impact of teacher training in higher education

Transfer of what is learned during professional training to actual professional practice is something which does not come about without effort (Baldwin & Ford, 1988; Thompson, Brooks & Lizarraga, 2003). Showers and Joyce (1996) stress that transfer does not occur automatically in the context of teacher training either: teachers frequently fail to implement what they have learned during a training program in their day-to-day teaching. Trainers should explicitly take into account the factors that explain why the supposed effects of training sometimes fail to emerge. Baldwin and Ford identify three categories of characteristics which influence transfer: characteristics of the training; characteristics of the participants; and characteristics of the context in which the subject is working. Joyce and Showers (1988) add a fourth category, namely: characteristics of the students. However, since students are part of a teacher's working context, they can be considered as coming under Baldwin and Ford's final category. Stes et al. (2007a) explored the various characteristics which course attenders taking part in a one-year training program for novice faculty members perceived as of overriding importance for the transfer of what they had learned. Factors perceived as constraining impact were mostly of a contextual nature, while the contextual aspect 'students' was perceived as an important promoting factor as well. The obtained results are in line with earlier research (Clarke & Hollingsworth, 2002; Trowler & Cooper, 2002) stressing that the impact of training depends heavily on the working context in which the subject operates.

Conclusions based on our literature search

On the basis of our literature search we conclude that the results of earlier research into the impact of teacher training in higher education do not always point in the same direction. Almost all the studies we examined investigated the impact on instructors' teaching. When students were involved, individual differences between students were not taken into account. We found only one study – that of Gibbs and Coffey (2004) – which looked at the impact on students' approaches to study. In their research the control group consisted of teachers and students from universities other than those from which the experimental group was drawn. We are not aware of previous research into the impact of teacher training in higher education on students' learning outcomes and there is also an apparent dearth of studies taking account of influencing factors. Our present study sets out to address these limitations.

The present study

In line with earlier studies, our research investigates the impact of instructional training for teachers in higher education on students' perceptions of teaching as well as on their learning approach and goes further by also examining the impact on students' learning outcomes. We conducted a multi-level analysis on the data, thereby allowing us to work with non-aggregated data and to take account of individual differences between students. Unlike in the case of the study by Gibbs and Coffey (2004), our control group of teachers all work at the same institution as the teachers who participated in the training program. Thus, if differences exist between the control and experimental groups, these cannot result from one group being in a different institutional context. Certain factors which can influence the impact of teacher training are also examined.

Research questions and hypotheses

Our central research questions and hypotheses are:

1. *Does a one-year training program for beginning teachers in higher education have an impact on their teaching, as perceived by students?*

Since most earlier research (Coffey & Gibbs, 2001; Gibbs & Coffey, 2004; Ho et al., 2001) found a positive impact of training on teaching behavior as rated by students, we hypothesized that, as a result of having received training, the teachers in our experimental group would obtain better ratings on their teaching practices at the post-test (after the training had been completed) than at the pre-test. In other words, we hypothesized that the difference scores (differences between post-test scores and pre-test scores) would be higher for the experimental teachers than for the teachers of the control group. We further hypothesized that at the end of the training program the teaching practices of the experimental group would be perceived as better by students, in comparison to the teaching practices of the instructors in the control group.

2. *Does a one-year training program for beginning teachers in higher education have an impact on students' study approach?*

On the basis of research by Gibbs and Coffey (2004) we hypothesized that, given that their teachers had undergone training, students of the experimental group would exhibit deeper study approaches at the post-test in comparison to the pre-test. In other words, we hypothesized that the difference scores (differences between post-test scores and pre-test scores) would be higher for the experimental students than for their fellow students in the control group. We further hypothesized that, at the post-test, the experimental group would have deeper study approaches than their fellow students in the control group.

3. *Does a one-year training program for beginning teachers in higher education have an impact on students' learning outcomes?*

We are not aware of previous research into the impact of training on students' learning outcomes. However, a study by Lea et al. (2003) reveals that learning environments in higher education based on student-centered learning have the potential to improve students' learning outcomes. We, therefore, hypothesized

that a training program intended to make the teaching practice of the participants better geared to the concept of student-centered learning would have a positive impact on student learning outcomes. We expected that, given the training their teachers had undergone, students of the experimental group would specify greater learning outcomes at the post-test in comparison to the pre-test. In other words, we hypothesized that the difference scores (differences between post-test scores and pre-test scores) would be higher for the experimental students than for the control students. We further hypothesized that, at the post-test, the experimental group would perceive greater learning outcomes than their fellow students in the control group.

4. *Is the impact of a one-year training program for beginning teachers in higher education, with regard to their teaching as perceived by students, student's study approach and learning outcomes, dependent on student numbers and student level (first years versus non first years)?*

We are not aware of any previous research that specifically addresses this research question. Stes et al. (2007a) explored the characteristics which participants in a one-year training program perceived as of overriding importance for the transfer of what had been learned. Factors perceived as constraining this impact mostly concerned contextual aspects, while the contextual aspect 'students' was perceived as an important promoting factor as well. We thus hypothesized that the impact of a one-year training program for beginning teachers would be dependent on the context in which the teachers concerned were working, namely on the number of students and their level (first years versus non first years).

Our study examines the effects of an instructional development program for beginning teachers at the University of Antwerp. Before presenting the design and results of the study, we will provide an outline of the training program.

The one-year training program for beginning teachers at the University of Antwerp

At the University of Antwerp novice teachers (defined as instructors appointed as lecturers or assistant lecturers during the last 7 years) may participate in a one-year training program on a voluntary basis. The participants constitute a heterogeneous group in terms of academic discipline and are limited to 25 so as to promote active involvement. The aim of the program is to raise professional standards with regard to classroom practice and to gear instructors' teaching more closely to the concept of competence-based and student-centered learning. This is an approach to teaching which has been explicitly embraced by the university management since 2000 and which emphasizes the development of competencies rather than the mere accumulation of knowledge, the ultimate goal of which is to challenge students to apply the knowledge and skills they have acquired in solving problems in complex real-life situations. The instructional training program for new staff consists of 4 modules, covering the following topics:

- 1) Activating teaching methods: what, how and why?
- 2) Assessing students
- 3) The 'Blackboard' electronic learning environment

4) Curriculum development

Participants have the opportunity to experience competence-based and student-centered learning at first hand by adopting the role of 'students'. They are encouraged to apply their knowledge and skills in each module by the use of activating teaching methods during contact moments, such as discussions based on video or micro-teaching fragments, case studies, guided autonomous learning, together with group and individual assignments accompanied by peer feedback. Home assignments at the end of each module stimulate the transfer of what is learned during the course to the novice teacher's current professional practice. These consist of the following tasks:

- A description of a teaching and learning environment, an outline for an activating teaching method and an outline for written material intended to stimulate their students' learning processes (module 1)
- A draft design for a valid and reliable assessment format or for an alternative assessment format (module 2)
- The development of a discussion forum, learning trajectory or test in the Blackboard electronic learning environment (module 3)
- An outline of the participants' course in relation to the curriculum (module 4)

At the beginning of the program the novice instructors were asked to choose one course out of the various courses they taught to which all their assignments would then be related.

Attending the contact moments and carrying out the assignments involved 140 hours of study, spread over a period of one academic year. In order to receive a certificate at the end of the training program those taking part were required to meet two criteria: to have attended at least 75% of the contact moments and to have completed the assignments set at the end of each module to a given standard. Award of a certificate is evidence of the recipient's active engagement in the training program and its value is thus guaranteed, so that it can be taken into account if the holder is being considered for promotion.

Participants in the present study

23 teachers who had enrolled in the above teacher training program at the University of X were invited to take part in the study. We also asked them to nominate a colleague who was not enrolled in the program, but whose teaching situation was otherwise comparable to their own, in terms of class size, the level of expertise of the students they taught, academic discipline, academic status and amount of teaching experience to serve as control teacher on a voluntary basis. Two teachers were unable to find a 'matching' colleague willing to take part and the resulting gap in numbers was filled by the research team. To prevent problems of subjects dropping out we found a total of four extra control teachers. The matches and the additional recruits thus constituted our control group. Our sample of 50 teachers (23 in the experimental group and 27 in the control group) then chose a single course from the various courses for which they were responsible on which they wished to concentrate during the study. The teachers in the experimental group were asked to select the course to which their home assignments during the training program had been related. A number of participants dropped out of the project while we were still

in the process of gathering data (February 2005 – October 2006). One of the teachers in the experimental group did not complete the training course and a further two failed to complete it successfully (meaning that no certificate was awarded on their termination of the program). Five control teachers dropped out either because they no longer worked at the university when the survey was conducted (n=2) or because, in practice, they were no longer responsible for the course they had originally selected for the purposes of our research (n=3). This left us with a final experimental group of 20 and a control group of 22.

The students taking the 42 courses selected by the participating teachers were also involved in the study. We tried to secure maximum student response by distributing questionnaires at the start of the class, where possible, and by giving students participating in the study the chance to win cinema tickets. Even so, student response was not always equally high for every teacher and at each of the data collection moments (pre- and post-tests), a feature which was automatically taken into account in our multi-level analysis as described below. 1,022 students took part in the pre-test, 471 of whom were taught by a teacher enrolled in the forthcoming training program and who thus belonged to the experimental group of students. 1,001 other students participated at the moment of the post-test, with 441 being taught by a certificated teacher. For five of the 42 participating teachers we were only able to collect student data relating to the pre-test. In this way, useful data (both pre- and post-test) were gathered for 19 experimental and 18 control teachers, who were to constitute our final sample.

Table 1 gives an overview of how this final sample of 37 teachers was categorized into the context variables 'Number of students in the classroom' and 'Students' level of expertise' (first years or not). With regard to the number of students taught by the participants, most taught small (≤ 30 students) or medium sized (> 30 but ≤ 100 students) groups. The distributions into control and experimental groups are comparable. Of the 19 experimental teachers, six were teaching first year students. A comparable proportion was found in the control group (seven teachers out of 18).

[Insert Table 1 about here]

Table 2 shows the distribution of the participants into the teacher characteristics of teaching discipline, gender, academic status and amount of teaching experience. Applying Becher's (1989) categorization of teaching disciplines, we can say that most teachers belonged to the 'hard' disciplines (e.g. chemistry or medicine). The 'soft' disciplines (e.g. history or education) are less well represented, especially in the experimental group. 26 of the total of 37 participants were male. Since only 19% of the teachers in the university concerned are female, it is quite normal that female teachers should be less well represented in our study. In the control group there were only a few females. With regard to academic status, most participants in our study were senior lecturers or professors. Assistant lecturers with a PhD are less well represented, especially in the control group. Of the 19 experimental teachers, 9 had five or less years of teaching experience (only years with a minimum teaching load of 15 hours were taken into account). In the control group the proportion of teachers with limited experience (≤ 5 years) was 5 out of 18.

[Insert Table 2 about here]

Instruments

Data were gathered by way of three questionnaires: one asked students about their perceptions of how the instructor taught the course; a second one questioned students about their study approach for the course concerned; and a third elicited information about students' learning outcomes for the course. Information about students' age, gender and the overall end mark obtained for the previous academic year was secured from the student administration office. In the case of first year students, who had, by definition, been at another institution the previous academic year (in most cases a secondary school) information about their overall end mark for that year was not taken into account due to problems of comparability, marks obtained in a different institution being difficult to compare in the educational context in which our study was conducted.

For each questionnaire we used a confirmatory factor analysis (CFA) to verify whether the model developed by the authors of the questionnaire could be validated. Where this proved not to be the case, an exploratory factor analysis (EFA) was conducted on a randomly selected sample comprising half the data. A confirmatory factor analysis could then be carried out on the model resulting from the EFA using the other half of the data. A brief description of each questionnaire and its validation is given below.

The questionnaire asking students about their perceptions of teacher's teaching was a translated version of part 2 of Entwistle's Experiences of Teaching and Learning Questionnaire (ETLQ), consisting of 40 items (Entwistle, 2005). It has six sub-scales, each representing a specific aspect of teaching: 'Clear aims, organization, alignment and integration'; 'Teaching for understanding and student choice'; 'Staff support and enthusiasm'; 'Assessment for understanding and feedback'; 'Interest, enjoyment and perceived relevance'; and 'Support from other students'. Themes such as making the aims of the course explicit, giving attention to the alignment and integration of aims, teaching and assessment, making both the teaching and the assessment as oriented towards conceptual understanding as possible and the importance of feedback as a tool for learning are explicitly addressed during the training program described above. Other aspects of teaching such as overall organization, giving students opportunities to choose, being supportive and enthusiastic as a teacher, facilitating peer support, capturing the interest of students and stimulating them to enjoy studying and to perceive the relevance of what is being learned, receive less attention during the training program. However, we decided not to remove the items/scales referring to these aspects from the original questionnaire, as they are all in line with the teaching concept of competence-based and student-centered learning which the training course seeks to promote. The student participants were thus presented with all 40 items, which were scored on a 5 point Likert scale ranging from 'was seldom or never true for me' to 'was always or nearly always true for me'. Students were allowed to give the response 'not applicable to this course'. A CFA was used to verify whether the six factor model as described above and elaborated by the authors of the ETLQ (Entwistle, 2005) could be validated. The results indicated that the data set fitted the model insufficiently (GFI=.81, AGFI=.79, CFI=.80, RMSEA=.07¹). A sufficient fit value for GFI, AGFI and CFI is a minimum of .90 (Hoyle, 1995), while for RMSEA a value of less than .08 is sufficient (Browne & Cudeck, 1993). Even

including error covariances we still could not obtain a sufficient fit. Personal e-mail communication with the authors of the questionnaire (January 16, 2006) revealed that in the final analyses of their ETLQ-data they had used a reduced set of items (n=25) and slightly different scales (n=8). We therefore conducted a CFA for this eight factor model and, after including five error covariances - all understandable with respect to the content of the items involved - the results indicated that the data fitted the model fairly well (GFI=.91, AGFI=.89, CFI=.91, RMSEA=.06). Acceptable Cronbach's alpha values were found for all scales: 'Aims and congruence' (Cronbach's alpha=.70); 'Teaching for understanding' (Cronbach's alpha=.73); 'Set work and feedback' (Cronbach's alpha=.76); 'Assessing understanding' (Cronbach's alpha=.63); 'Staff enthusiasm and support' (Cronbach's alpha=.67); 'Student support' (Cronbach's alpha=.72); and 'Interest and enjoyment' (Cronbach's alpha=.87). Only the scale 'Student choice' turned out not to be reliable (Cronbach's alpha=.44) and could therefore not be used for further analysis. As mentioned above, this scale is not related to a central subject within the program, so we did not consider this to be a problem. For further analysis of the data on students' perceptions of the teaching we used the eight factor model described above, but without the unreliable scale 'Student choice'. The scales are listed in Table 3, together with a typical questionnaire item.

To investigate students' study approaches we used the Dutch translation by Gijbels, Van de Watering, Dochy and Van den Bossche (2005) of Biggs, Kember and Leung's (2001) Revised Two Factor Study Process Questionnaire (R-SPQ-2F), as a point of departure, but with one important modification, namely that all items were formulated in the past tense and made course-specific. For example, an original item such as 'Studying gives me a deep feeling of personal satisfaction from time to time' became 'Studying for this course gave me a deep feeling of personal satisfaction from time to time'. Gijbels et al. (2005) investigated the psychometric properties and applicability of the R-SPQ-2F in the Dutch context. Their results indicated good reliability coefficients and goodness of fit for the two factor solution with deep and surface learning approaches. As a first step in the validation we conducted a CFA on the two factor solution as reported in Gijbels et al. (2005). The fit indices for this model showed that the two factor solution was a poor fit with the data (GFI=.86, AGFI=.82, CFI=.80, RMSEA=.09). As a subsequent step in the validation of the questionnaire we thus conducted an EFA. As we wanted to perform a CFA on the model resulting from the EFA, the data of only (a randomly selected) half of the students who had filled in all items of the questionnaire (n=967) was taken into account. The EFA resulted in a five factor model with scales which can be described as: 'Studying with as little effort as possible'; 'Learning by heart'; 'Studying is interesting'; 'Spending extra time on studying,' and 'Self-regulated learning'. Although the EFA did not lead to a two factor solution we observed a clear link between the scales 'Studying with as little effort as possible' and 'Learning by heart' and a factor as 'Surface approach'. For the other three scales the link was seen with a factor as 'Deep approach'. Thus, since earlier studies with the SPQ²/R-SPQ-2F, including cross-cultural studies (e.g. Gijbels et al., 2005; Kember & Leung, 1998; Snelgrove & Slater, 2003; Watkins & Regmi, 1996; Zhang, 2000), indicated that a model with two factors had a good fit, we decided to conduct a CFA on a model with two heading scales ('Deep approach', 'Surface approach') and the five sub-scales as described above, whereby the scales

‘Studying with as little effort as possible’ and ‘Learning by heart’ were linked to the heading scale ‘Surface approach’; and the scales ‘Studying is interesting’, ‘Spending extra time on studying’ and ‘Self-regulated learning’ were linked to the heading scale ‘Deep approach’. The results of a CFA on this model (using the data from the other half of students, $n=1007$) revealed that the data fitted the model well ($GFI=.93$, $AGFI=.91$, $CFI=.92$, $RMSEA=.06$). Acceptable Cronbach’s alpha values were found for the two heading scales: ‘Surface approach’ (Cronbach’s $\alpha=.76$) and ‘Deep approach’ (Cronbach’s $\alpha=.71$). We decided to delete one item (‘The course was not very interesting, so I only made a minimal effort’) from both heading scales as its exclusion raised the Cronbach’s alphas to $.84$ for the scale ‘Surface approach’ and to $.82$ for the scale ‘Deep approach’. The Cronbach’s alphas for the sub-scales were also good: ‘Studying with as little effort as possible’ (Cronbach’s $\alpha=.82$); ‘Learning by heart’ (Cronbach’s $\alpha=.73$); ‘Studying is interesting’ (Cronbach’s $\alpha=.81$); and ‘Spending extra time on studying’ (Cronbach’s $\alpha=.78$). Only the sub-scale ‘Self-regulated learning’ was revealed to be unreliable (Cronbach’s $\alpha=.44$) and could not be used for further analysis. Thus, for further analysis of the data relating to students’ study approach we used a model with two factors and four sub-factors. Table 3 gives the sub-factors together with a typical questionnaire item, by way of illustration.

The questionnaire asking students about their learning outcomes consisted of a translated version of part 4 of the ETLQ-questionnaire (Entwistle, 2005) extended by the addition of items intended to measure affective and psychomotor learning outcomes. From our personal e-mail communication with the authors of the ETLQ-questionnaire (January 16, 2006) we learned that in the final analyses of their data they used the following scales for part 4 of the questionnaire: ‘Knowledge and subject-specific skills’, ‘Generic skills’ and ‘Information skills’. These three scales and the related nine items were all concerned with cognitive learning outcomes. Guskey (2000) stresses the importance of considering not only students’ cognitive outcomes, but also their affective learning outcomes (changes in interest, enjoyment, self-confidence and responsibility) and their psychomotor learning outcomes (changes in study behavior) when evaluating teachers’ professional development. For this reason we constructed four items which measured affective learning outcomes as well as a further four items measuring psychomotor learning outcomes and added them to part 4 of the ETLQ-questionnaire. As a first step in the validation of the questionnaire we conducted a CFA on the model that we expected on the basis of how the questionnaire was constructed, namely a model with three scales: ‘Cognitive learning outcomes’ (consisting of the nine items of part 4 of the ETLQ); ‘Affective learning outcomes’ (consisting of four items constructed by us); and ‘Psychomotor learning outcomes’ (consisting of four items constructed by us). The results indicated that the data set had an insufficient fit with the model ($GFI=.79$, $AGFI=.73$, $CFI=.72$, $RMSEA=.12$). As a subsequent step in the validation of the questionnaire we carried out an EFA on data from only (a randomly selected) half of the students who had filled in all items of the questionnaire ($n=972$). The EFA resulted in a four factor model with scales which can be described as: ‘Affective learning outcomes’; ‘Psychomotor learning outcomes’; ‘Generic and information skills’; and ‘Knowledge and subject-specific skills’. The three-fold division as described by Guskey (2002) can be recognized, with two separate scales relating to cognitive learning outcomes (‘Generic

and information skills' and 'Knowledge and subject-specific skills'). The scale 'Knowledge and subject-specific skills' was also used by Entwistle (personal e-mail communication, January 16, 2006); although he divided the scale 'Generic and information skills' into two. The results of a CFA on our four factor model (using the data of the other half of students, n=1010) revealed that the data fitted the model well (GFI=.94, AGFI=.91, CFI=.93, RMSEA=.07). Acceptable Cronbach's alpha values were found for all scales: 'Affective learning outcomes' (Cronbach's alpha=.80); 'Psychomotor learning outcomes' (Cronbach's alpha=.76); 'Generic and information skills' (Cronbach's alpha=.78); and 'Knowledge and subject-specific skills' (Cronbach's alpha=.74). We thus used the four factor model described above for further analysis of the data relating to students' learning outcomes. Table 3 lists each scale, together with a typical questionnaire item.

[Insert Table 3 about here]

Data analysis

Given the design of the study (i.e. that the same teachers and courses were considered at the post-moment as had been considered at the pre-test) the students involved in the post-test were not the same individuals as those of the pre-test (except for repeating students).

In the multi-level analysis as described below the extent to which the two groups of students (pre-test group and post-test group) are comparable with regard to gender, age and general academic level (in terms of global end mark obtained during the previous academic year) is taken into account. Comparability with regard to these variables is an important consideration as these can influence students' perceptions of the learning environment, their study approach and learning outcomes (Richardson, 2006). Before reporting the results of the multi-level analysis we will briefly describe the mean pre-test scale scores with regard to the (standardized) dependent variables for both the experimental and the control group of teachers. All student data were likewise aggregated. The extent to which the two instructor groups differ with regard to their teaching as perceived by students, their students' study approaches and learning outcomes at the moment of the pre-test, is considered in the multi-level analysis. As a first step in our multi-level analysis³ we estimated a 'null model', with no explanatory variables. This gave us information about teachers' mean post-test score and mean difference score (i.e. the difference between the post-test and pre-test score) on the dependent variable concerned. It also enables us to distinguish the variance in the dependent variable at teacher level from the variance at student level (by calculating the intra-class correlation – i.e. the percentage of variance at teacher level in relation to the total variance at teacher and student level together). Only if scores differ in the teachers' group (post-test and/or difference) does it make sense to examine whether training has an effect.

The multivariate model contains two levels: the student level (level 1) and the teacher level (level 2), whereby the students are nested in teachers. The following formula shows how the 'null model' is put together.

$$Y_{ij} = \beta_{0j} * \text{Constant} + \beta_{1j} * \text{Difference score}_{ij} + \text{Constant} * v_{0j} + \text{Difference score} * v_{1j} + \varepsilon_{2ij} * \text{Pre-test score}_{ij} + \varepsilon_{3ij} * \text{Post-test score}_{ij}$$

Y_{ij} stands for the score on dependent variable Y of student i belonging to teacher j . In the fixed part of the formula ($\beta_{0j} * \text{Constant} + \beta_{1j} * \text{Difference score}_{ij}$), Constant has the value 1 for all students, $\text{Difference score}_{ij}$ is a dummy variable indicating whether a student belongs to the pre-test group (value -1) or to the post-test group (value 0). Thus the estimated regression coefficients β_{0j} and β_{1j} are the mean post-test score and the mean difference score, respectively.

However, we were not interested in the mean scores as such, but rather in the differences in scores between teachers. We therefore added a random part ($\text{Constant} * v_{0j} + \text{Difference score} * v_{1j}$) to the formula whereby v_{0j} stands for the degree to which a teacher's post-test score deviates from the mean post-test score and v_{1j} stands for the degree to which a teacher's difference score deviates from the mean difference score. We also take into account the possibility of a covariance between v_{0j} and v_{1j} .

As we expected inter-individual differences in scores between students of the same teacher, we wanted to be able to include these in the model. We thus had to add two dummy variables indicating if a student belongs to the pre-test group or to the post-test group: $\text{Pre-test score}_{ij}$ (values 0 and 1 for a student belonging to the post-test and pre-test group, respectively) and $\text{Post-test score}_{ij}$ (values 0 and 1 for a student belonging to the pre-test and the post-test group, respectively). For these two dummy variables we added a random effect to the model: ε_{2ij} and ε_{3ij} which stands for the inter-individual differences between students at the pre-test and at the post-test, respectively.

A second model, which we call the 'gross model', reported the effect of belonging to the experimental or the control group of teachers with regard to both the post-test score and the difference score on the dependent variable concerned. The impact of participation in the one-year teacher training is examined, regardless of the possible impact on the dependent variable, by way of (other) teacher characteristics (teaching discipline, gender, academic status, teaching experience); context characteristics (number of students in the course) and student characteristics⁴ (gender, age, academic level⁵). The results of the gross model provide an insight into the effects of teacher training as they occur in day-to-day teaching practice. However, the effect of participation in training can be overruled by the effects of other teacher characteristics, the context or the students.

A third model, the 'net model', is of a more hypothetical nature: it makes us think about the effects of teacher training when considering a hypothetical mean teacher, mean context and mean student. The model examines the effect of the teacher, context and student characteristics as mentioned above, on the dependent variable. We were particularly interested in the net impact of teacher training, namely the impact after verification for possible effects of other components in the model.

Finally, two 'interaction models' were built in which we investigated the impact of teacher training in interaction with two characteristics, namely the number of students and their academic level (first years or non first years). In the interaction models the same teacher, context and student characteristics were included as in the net model; only the effects of training in interaction with the students' level (first interaction model) and the number of students (second interaction model), respectively, were specifically added.

In the results section we describe the findings for the different models with respect to each of the dependent variables (students' perceptions of teaching; students' study approaches; and students' learning outcomes). We report the estimation and (in brackets) the standard error for each regression coefficient. If the value of the estimation is greater than or equal to double the value of the standard error, the effect is significant (Hox, 2002). Since the number of participating teachers in our study was rather limited we will consider not only statistically significant effects, but also results which are interesting from a practical point of view. Regression coefficients with an absolute value equal to or greater than .2, .5 and .8 will be considered as indicating a small, medium and large effect respectively (Cohen, 1988).

To aid interpretation and comparison between scores on different scales, all dependent variables were standardized.

Results

The mean scale scores of the experimental and control group at the pre-test

Table 4 shows the mean scale scores and (in brackets) the standard deviations with regard to the (standardized) dependent variables for the experimental and control group at the pre-test, calculated on the basis of aggregated student data. When we examine the scales concerning students' perceptions of teaching, we observe that, in general, the experimental teachers score higher than the control teachers on all scales. The same can be said with regard to the scales concerning student's learning outcomes as well as for the scale 'Deep study approach' and its two sub-scales ('Studying is interesting' and 'Spending extra time'). The opposite is true for the mean scores on the scale 'Surface study approach' and its sub-scales ('Studying with minimal effort' and 'Learning by heart'): on these scales the control group has the highest mean.

At the moment of the pre-test the experimental group of teachers and the control group are not comparable with regard to their teaching as perceived by their students, nor with respect to their students' study approach and learning outcomes. In general, the teaching of the instructors from the experimental group is perceived as better; their approach provides a greater stimulus to students to adopt a deep study approach rather than a surface one and results in greater (perceived) learning outcomes. We conclude that at the moment of the pre-test the experimental teachers meet an important goal of the training program to a greater extent than the teachers in the control group, namely being able to teach in a competence-based and student-centered way. However, looking at the standard deviations, we can see that there are quite strong inter-individual differences in both groups.

The fact that at the pre-test the experimental teachers differ from the control teachers with regard to their teaching as perceived by students, their students' study approaches and learning outcomes is taken into account in the different models used in our multi-level analysis by considering not only post-test scores but also difference scores. As the mean scale scores were calculated on the basis of aggregated student data, differences between the experimental group and the control group of teachers may also result from differences between the two groups of students (experimental and control groups) with regard to gender, age and academic level. The 'net model' of the multi-level analysis takes into

account the comparability of the two groups of students with regard to these characteristics. This model also controls for differences between the experimental and control teachers resulting from teacher characteristics such as discipline, gender, academic status, amount of teaching experience or from context characteristics (number of students).

[Insert Table 4 about here]

Effects of teacher training on students' perceptions of teaching

The output data for the null model (Table 5) shows that teachers differ from each other with respect to their scores on the scales measuring students' perceptions of teaching. With regard to the post-test scores, all variances at teacher level are significant. Variances in the difference scores at teacher level are smaller. They are significant for all scales except 'Student support'. However, differences in scale scores also result to a large extent from differences between students. Intra-class correlations (ICC) are situated between 18% ('Teaching for understanding') and 31% ('Set work and feedback'). The differences between students of the same teacher are larger than the differences between teachers for all scales. This indicates that, in our context, when using the ETLQ (part 2) as an instrument to measure students' perceptions, the same teaching can be perceived in a range of ways by different students. In fact, given our context, the ETLQ (part 2) is probably not the most appropriate instrument for providing a clear indication of instructors' teaching. It leads to strong differences in scale scores at student level making it relatively difficult to find an effect of teacher training on these scale scores at teacher level. The correlations between the variance at teacher level with regard to the post-test score and those with regard to the difference score are moderately positive for the scales 'Aims and congruence', 'Set work and feedback' and 'Assessing understanding'. This means that, in general, teachers scoring high on these scales at the post-moment also made the most gains in score from pre-test to post-test.

The output data relating to the gross model are reported in Table 6. We see that there is no significant effect of participation in teacher training on the post-test scores or on the difference scores for the scales concerning students' perceptions of teaching. With regard to the post-test scores, we note a small negative effect of participation in training on the scale 'Interest and enjoyment' (-.22 SD). For the difference scores, there is a small positive effect of training on the scale 'Set work and feedback' (.21).

The full output data for the net model are presented in Appendices 1a (data relating to teacher characteristics) and 1b (data relating to context and student characteristics). The output data relating to the net impact of teacher training are shown in Table 6. Only the results regarding this impact are reported in the body of the present article. The statistically significant net impact of teacher training with regard to students' perceptions of teaching is limited to the scale 'Teaching for understanding' and – interestingly – is negative. Teachers who participated in our teacher training as described above, score .28 SD lower on this scale than their fellow teachers in the control group. Small negative effects are noticed with regard to the post-test score on the scale 'Interest and enjoyment' (-.33 SD).

Table 7 shows the output data relating to the effect of teacher training in interaction with the student level (first years versus non first years). When we compare the data for this interaction model with the data for the net model, we observe that the net and negative impact of teacher training with regard to the post-test score on the scale 'Teaching for understanding' as shown in Table 6 diminishes if the distinction is made between teachers with first year students and those teaching non first year students. The impact is no longer statistically significant; however, there is still a small practical effect (-.29 SD) for teachers with non first year students. The impact on the post-test score for the scale 'Interest and enjoyment' (-.46 SD) becomes statistically significant for those teaching non first years. The negative net effect with regard to the post-test score on 'Interest and enjoyment' becomes positive when only teachers with first year students are considered (.44 SD). The output data for the interaction model as shown in Table 7 show that the post-test and/or the difference scores on a considerable number of other scales also indicate small practical effects when distinguishing those teaching first years from those teaching non first year students.

Table 8 shows the output data relating to the effect of teacher training in interaction with the number of students. When we compare these data with the data for the net model, we observe that the net and negative impact of teacher training with regard to the post-test scores on the scale 'Teaching for understanding' as shown in Table 6 changes if the distinction is made between instructors teaching a small (≤ 30), medium-sized (>30 but ≤ 100) or large number (>100) of students. The impact is no longer statistically significant; there is a small practical and negative effect for instructors teaching a small group (-.44 SD), whereas there is a small positive effect for teachers teaching a medium number of students (.28 SD). The negative effect noticed in the net model with regard to the post-test scores on the scale 'Interest and enjoyment' is still present in the interaction model when taking into account instructors teaching small groups (-.50 SD) but, in the interaction model, becomes positive for the other teachers (.25 SD and .49 SD for those teaching a medium and large number of students, respectively). The output data as shown in Table 8 shows that the post-test and/or the difference scores for a fair number of other scales also indicate small to large practical effects when distinguishing those teaching a small, medium and large number of students. There is a large and significant effect with regard to the difference scores on 'Set work and feedback' for teachers with a group of medium size. A medium and significant effect is noticed with regard to the difference scores on 'Student support' for teachers with large groups.

Effects of teacher training on students' study approaches

The output data for the null model (Table 5) indicate that teachers differ from each other with respect to their students' study approaches. With regard to the post-test scores, as well as the difference scores, the variances at teacher level are significant for all scales. Differences in scale scores also result to a large extent from differences between students. Intra-class correlations (ICC) are situated between 19% ('Studying with minimal effort' and 'Learning by heart') and 26% ('Spending extra time'). The differences between students of the same teacher are larger than the differences between teachers for all scales and sub-scales, which suggests that it will be quite difficult to find an effect of teacher training on these scale and sub-scale scores at teacher level. The correlations between the

variance at teacher level with regard to the post-test score and those with regard to the difference score are moderately to strongly positive (except for the scale ‘Studying is interesting’). This means that, in general, teachers with students who score highly on these scales at the post-moment, have students who also made the greatest gain in score from pre-test to post-test.

The output data relating to the gross model are reported in Table 6. We see that there is no significant effect of participation in teacher training on the post-test scores, nor on the difference scores for the scales relating to students’ study approaches. With regard to the post-test scores, we note a small negative effect of participation in training on the scale ‘Studying is interesting’ (-.24 SD).

When we look at the output data relating to the net impact (Table 6) we see a significant net and negative impact of teacher training (-.31 SD) for the post-test scores on the scale ‘Studying is interesting’. Small effects of training can be found with regard to the post-test scores on the scale ‘Surface approach’ (.28 SD) and its sub-scales ‘Studying with minimal effort’ (.28 SD) and ‘Learning by heart’ (.22 SD). Small negative effects are noticed with regard to the difference scores on the scale ‘Studying is interesting’ (-.21 SD).

Table 7 shows the output data relating to the effect of teacher training in interaction with student level (first years versus non first years). When we compare the data for this interaction model with the data for the net model, we notice that the negative net effect of teacher training for the post-test score on the scale ‘Studying is interesting’ remains statistically significant when considering teachers with non first year students; the effect becomes even stronger (-.46 SD). Moreover, the impact on the scale ‘Surface approach’ (.44 SD) and its sub-scale ‘Studying with minimal effort’ (.41 SD) becomes statistically significant for those teaching non first years. As in the net model small practical effects of training can be found with regard to the post-test scores on the scale ‘Learning by heart’. However, the effect is positive (.36 SD) when it concerns teachers with non first year students, and negative for those teaching first year students (-.40 SD). The small negative net effect with regard to the difference scores on the scale ‘Studying is interesting’ is also noticeable in the interaction model, for teachers with non first year students as well as for teachers with first year students. The output data for the interaction model as shown in Table 7 indicate that the post-test and/or the difference scores for a considerable number of other scales concerning students’ study approaches also indicate small practical effects when distinguishing those teaching first years from those teaching non first year students. With regard to the difference scores there is a medium negative effect on ‘Deep approach’ for teachers with first year students.

Table 8 shows the output data concerning the effect of teacher training in interaction with the number of students. When we compare these data with the data for the net model, we note that the negative net effect of teacher training for the post-test score on the scale ‘Studying is interesting’ remains statistically significant when considering those teaching small groups (≤ 30 students); the effect for this group of teachers is even stronger (-.69 SD). Furthermore, the impact on the scale ‘Deep approach’ (-.57 SD) becomes statistically significant for teachers with small groups. As in the net model small effects of training which are interesting from a practical point of view can be found with regard to the post-test scores on the scale ‘Learning by heart’; but not for teachers with medium-sized groups (>30 but ≤ 100 students). The effect is positive (.29 SD) with regard to

teachers of small groups, and negative for teachers of large groups (>100 students) (-.34 SD). The negative net effect with regard to the difference score on the scale 'Studying is interesting' is also noticeable in the interaction model, but not for teachers with a medium number of students. The output data given in Table 8 indicate that the post-test and/or difference scores for several other scales relating to students' study approaches indicate small to medium practical effects.

Effects of teacher training on students' learning outcomes

The output data for the null model (Table 5) suggest that teachers differ from each other with respect to their students' learning outcomes. With regard to the post-test scores, the variances at teacher level are significant for all scales. With regard to the difference scores, we only see a significant effect for the scale 'Knowledge and subject-specific skills'. Differences in scale scores also result, to a large extent, from differences between students. Intra-class correlations (ICC) are situated between 11% ('Knowledge and subject-specific skills') and 25% ('Generic and information skills'). The differences between students of the same teacher are larger than the differences between teachers and this applies to all the scales concerning student learning outcomes, which suggests that it will be quite difficult to find an effect of teacher training on the scale scores at teacher level. For the scale 'Psychomotor skills' the correlation between the variance at teacher level with regard to the post-test score and those with regard to the difference score is moderately negative. Thus teachers with students scoring high on this dependent variable at the post-test moment have students who made the least gains from pre-test to post-test. The output data relating to the gross model are reported in Table 6. We note that there is no significant effect of participation in teacher training on the post-test scores, nor on the difference scores for the scales relating to student learning outcomes. Regarding the post-test as well as the difference scores, we observe small negative effects of participation in training on the scale 'Affective learning outcomes' (-.26 SD respectively -.21 SD). The statistically significant net impact of teacher training on students' learning outcomes (Table 6) is limited to the scale 'Knowledge and subject-specific skills', with students of teachers who participated in training, scoring .26 SD lower on the post-test. Small negative effects are noticed with regard to the post-test scores on the scales 'Affective learning outcomes' (-.27 SD) and 'Generic and information skills' (-.24 SD), as well as with regard to the difference score on the scale 'Affective learning outcomes' (-.24 SD). Table 7 shows the output data relating to the effect of teacher training in interaction with student level (first years versus non first years). When we compare the data for this interaction model with the data for the net model, we see that the negative net effect of teacher training for the post-test score on the scale 'Knowledge and subject-specific skills' remains statistically significant in our interaction model when considering teachers with non first year students; the effect becomes even stronger (-.37 SD). The small negative effects noticed in the net model with regard to the post-test scores on the scales 'Affective learning outcomes' and 'Generic and information skills' are still present in the interaction model when taking into account teachers of non first years (-.31 SD and -.22 SD, respectively) but are not observed for teachers teaching first year students (.15 SD and -.07 SD, respectively). In the interaction model the small negative net effect with regard to the difference score on the scale 'Affective learning outcomes' is only present

for teachers of non first year students (-.21 SD). The output data for the interaction model as shown in Table 7 reveal that some other post-test and difference scale scores also indicate small practical effects when distinguishing teachers with first year students from those teaching non first year students.

Table 8 shows the output data relating to the effect of teacher training in interaction with the number of students. When we compare these data with the data for the net model, we see that the net and negative impact of teacher training with regard to the post-test score on the scale 'Knowledge and subject-specific skills' diminishes. The impact is no longer statistically significant; there is a small negative effect for teachers teaching small groups (≤ 30 students) (-.49 SD); for teachers with medium-sized (>30 but ≤ 100) or large groups (>100) there are small positive effects (.38 SD and .23 SD, respectively). The impact on the post-test score for the scale 'Psychomotor learning outcomes' becomes statistically significant for teachers with large groups. The negative effects noticed in the net model with regard to the post-test scores on the scales 'Affective learning outcomes' and 'Generic and information skills' are still present in the interaction model when taking into account teachers of small groups (-.52 SD and -.49 SD, respectively) but in the interaction model they become positive for teachers with medium-sized (.35 SD or .37 SD, respectively) or large groups (.43 SD and .35 SD, respectively). The small negative net effect with regard to the difference scores on the scale 'Affective learning outcomes' disappears in the interaction model. The output data as shown in Table 8 suggest that some other post-test and difference scale scores also indicate small practical effects when distinguishing teachers on the basis of the number of students they have to teach.

[Insert Tables 5 to 8 about here]

Conclusions and discussion

In this study we looked at whether there are differences in teaching as perceived by students between teachers in higher education who have participated in instructional training and those who have not. We also explored the effects of instructional training on students' study approach and students' learning outcomes. A multi-level analysis was conducted on the quantitative data gathered at both a pre-test and a post-moment, involving more than 1,000 students at each point. We worked with non-aggregated data and individual differences between students were taken into account.

Before conducting our multi-level analysis we considered the mean scale scores on the dependent variables at the pre-test. The experimental group of teachers and the control group were revealed to be non-comparable with regard to their teaching as perceived by their students, their students' study approach and learning outcomes. In general, the teaching of the instructors in the experimental group was perceived as better; provided a greater stimulation to students to adopt a deep study approach rather than a surface approach and resulted in greater (perceived) learning outcomes. Thus, before the start of the training program, the experimental teachers met an important training goal to a greater extent than the teachers of the control group, namely that of being able to teach in a competence-based and student-centered way. This may have influenced the overall impact of the training: teachers who already score highly on the dependent variables will have more difficulties in improving still further than teachers with low scores. There were

fairly marked inter-individual differences, however. We cannot say that all those in the experimental group were teaching in a more competence-based and student-centered manner than those in the control group at the moment of the pre-test.

Our first model of the multi-level analysis ('the null model') showed that the teachers involved in our study (in both the experimental and the control group) differ from each other with regard to their teaching as perceived by students, their students' study approach and their students' learning outcomes, which means that exploring the effect of training is worthwhile. However, differences in scale scores also resulted to a large extent from differences between students. Differences between students of the same teacher were larger than the differences between teachers for all the dependent variables concerned. Given these strong differences in scale scores at student level, it is quite difficult to find an effect of teacher training on these scale scores at teacher level.

When taking into account the results relating to the gross and net models, we thus have to give a qualified answer to our first research question: *'Does a one-year training program for beginning teachers in higher education have an impact on their teaching, as perceived by students?'* Whereas the gross model revealed no significant effects, the net model indicated one effect, namely a negative effect on the scale 'Teaching for understanding'. A small effect which is interesting from a practical point of view was found on the teaching aspect 'Interest and enjoyment', in both the gross and the net models. Again, the effect was negative. Our hypothesis that at the end of training the teaching of the experimental group would be perceived by students as better in comparison to that of the teachers in the control group is not confirmed. In fact, the contrary is true: with respect to the aspects 'Teaching for understanding' and 'Interest and enjoyment', the control teachers are rated as better. When considering the difference scores, a small positive (although not statistically significant) effect was found in the gross model with regard to the scale 'Set work and feedback'. Thus, our hypothesis that teachers in our experimental group receive better ratings on their teaching behavior at the post-test (once training has come to an end) than at the pre-test, as a result of their having undergone training, is only confirmed with regard to this single aspect of teaching. Our findings contradict earlier research results from Ho et al. (2001), Coffey and Gibbs (2001) and Gibbs and Coffey (2004) which indicated a positive impact of training on teaching behavior as rated by students. This impact was not apparent in our study. It is possible that a longer time period is required in order to find an impact of training on teaching behavior. Earlier studies by Gibbs and Coffey (2004), Stes et al. (2007b) and Postareff et al. (2007) stressed that the shift to a (more) conceptual change/student-focused approach resulting from training is a slow process. Teachers taking part in instructional training programs may, on completion of their training, wish and try to adopt a more student-centered approach, but may encounter difficulties in implementing this (no impact on teaching behavior) or in doing so correctly (negative impact on teaching behavior). Follow-up research is needed to investigate whether there is indeed a positive impact of training on teaching behavior after a longer period of time has elapsed. On the other hand, however, we must beware of jumping to the conclusion that training has no impact on teaching behavior at all. In contradiction to earlier studies by, for example, Rust (1998) and Stes et al. (2007a) which relied only on teachers' self-reporting, we used student perceptions to investigate teaching behavior. However, students' expectations with regard to higher education teaching and their preferences with

respect to teaching were not taken into account and this may have influenced our results. That effects were not perceived or appreciated by the students does not automatically mean that they did not occur. There is often a gap between the objective characteristics of a teaching and learning environment and student perceptions of this environment (Ramsden, 1997). That said, however, student perceptions certainly have to be taken into account as they are crucial to the success of teaching innovations (Ramsden, 1997). Entwistle (1998) stresses that it is not the teaching as such, but rather the way students perceive teaching, that affects student learning.

This brings us to our second research question: *‘Does a one-year training program for beginning teachers in higher education have an impact on students’ study approach?’* Whereas the gross model revealed no significant effects, the net model indicated one effect, namely a negative effect on the scale ‘Studying is interesting’. In the gross model a small effect which is interesting from a practical point of view and (again) negative was found with respect to this scale. A small positive impact was observed in the net model with regard to the scale ‘Surface approach’ and its sub-scales ‘Studying with minimal effort’ and ‘Learning by heart’. Our hypothesis that, at the post-test, students in the experimental group would have deeper study approaches in comparison to their fellow students in the control group is not confirmed. In fact, with respect to the aspect ‘Studying is interesting’, the control students fare significantly better. When considering the difference scores, a negative (although not statistically significant) effect was found in relation to the scale ‘Studying is interesting’. Thus our hypothesis that, as a result of the training their teachers had undergone, experimental students would have a deeper study approach at the post-test than at the pre-test moment is not confirmed. Our results, thus, contradict earlier findings from Gibbs and Coffey (2004), but are in line with the idea advanced by Entwistle (1998) and mentioned above, namely that students’ perceptions of teaching, affect their learning. Given that, in our study, students’ perceptions seemed not to have been affected (see the results relating to our first research question), it might be regarded as no surprise that there should be no impact on students’ study approach either.

Our third research question was related to students’ learning outcomes: *Does a one-year training program for beginning teachers in higher education have an impact on students’ learning outcomes?* While the gross model revealed no significant effects, the net model indicated one effect, namely a negative effect on the scale ‘Knowledge and subject-specific skills’. Small effects which are interesting from a practical point of view were found on ‘Affective learning outcomes’ (in both the gross and the net models) and ‘Generic and information skills’ (in the net model only). Again the effects were negative. Our expectation that students of the experimental group would state greater learning outcomes at the post-test in comparison to the control group was not confirmed. On the contrary, control group students mention more knowledge and subject-specific skills, affective learning outcomes and generic and information skills than the experimental students. When considering the difference scores, a small negative (although not statistically significant) effect was found with regard to ‘Affective learning outcomes’. Thus our expectation that, as a result of the training their teachers had undergone, experimental students would report greater learning outcomes at the post-test than at the pre-test was not confirmed. We are not aware of previous research into the impact of teacher training in higher education on students’ learning outcomes. However, a study by

Lea et al. (2003) revealed that learning environments in higher education which are based on student-centered learning have the potential to improve student learning outcomes. Our results clearly show that this potential was not realized, although further research is needed to investigate the reasons for this. It is possible that, despite the training they received, participants did not actually gear their teaching practice to the concept of student-centered learning (we saw no impact on teaching as perceived by students - see research question 1). It might also be that this was simply not perceived by students and therefore did not affect their study approach (see research question 2) and there was similarly no increase in learning outcomes.

Our final research question concerned the impact of training in interaction with context characteristics: *Is the impact of a one-year training program for beginning teachers in higher education, with regard to their teaching as perceived by students, student's study approach and learning outcomes, dependent on student numbers and student level (first years versus non first years)?* Our first interaction model revealed a differential impact of training for teachers with first year students and those teaching non first years. In general terms, the impact is more positive for those teaching first years with regard to (most aspects of) their teaching as perceived by the students as well as with regard to students' study approach and learning outcomes; however, the differences in scale scores between experimental teachers with first year students and control teachers with first years are not significant. A second interaction model revealed a differential impact of training depending on the size of the group of students taught. Overall, the impact is more positive for those teaching larger groups (medium-sized or large) with regard to (most aspects of) their teaching as perceived by the students, as well as with regard to students' study approach and learning outcomes. However, there are very few significant differences in scale scores between experimental teachers with medium-sized or large groups and control teachers with medium-sized or large groups.

When taking into account all the results obtained in relation to our four research questions, we conclude that the effects of an instructional development program for beginning teachers at the University of Antwerp were limited when the impact on teaching as perceived by the students, on students' study approaches and students' learning outcomes are considered. There may be several reasons for these findings. First, all our respondents, including the control teachers, participated voluntarily. This may mean that the control group teachers were more interested in education and paid more attention to good teaching than the average university teacher, which may have influenced the results. Differences between experimental and control group teachers might have been more noticeable if we had been working with a control group of randomly selected teachers instead of a volunteer control group. However, even when a design of this sort is used it is still difficult to prevent the so-called 'Hawthorne effect' (Shayer, 1992): the mere fact of being involved in the impact study (as a member of the control group) provides an extra stimulus for individuals to think about their teaching and can be the starting point towards a modification of their teaching. The Hawthorne effect may thus result in the impact of training being underestimated and further research into this is clearly needed. The control teachers did not participate in the instructional development program but may have improved as teachers by other means such as by reading books about competence-based and student-centered learning or by asking their

students to give them feedback on their teaching. The university also offers shorter instructional training activities in addition to the one-year training program and these additional forms of professional development also need to be taken into account in future research.

Two interaction models examined the interfering impact of the variables ‘student level’ (first years or not) and ‘student numbers’. Other possible intervening variables such as teacher’s satisfaction with the quality of the training, their motivation for teaching or their perception of the degree to which teaching is valued in their institution (in comparison to research) can also influence the impact of training and should be examined in future research.

We have to be careful in generalizing our findings. The number of teachers for whom data were gathered was relatively small, which increases the chance of finding no effects in a quantitative study, while in practice there are differences between teachers who followed the training program and those who did not. We tried to overcome this problem as far as possible by considering not only statistically significant effects, but also results which are interesting from a practical point of view, but this still remains an attention point. Finally, our study examined the impact of a specific training program in a specific university context. Our findings cannot be unthinkingly generalized to other kinds of instructional training and/or to other settings.

Our study shows that instructional training for teachers in higher education does not automatically result in effects on students’ perceptions of teaching, their study approach or learning outcomes. Although the training program examined in this study explicitly tried to support participants in making their teaching more geared to the concept of competence-based and student-centered teaching, it seems that (still) more attention needs to be paid to the implementation of the different aspects of competence-based and student-centered teaching in daily teaching practice, to ways of inhibiting surface study approaches and stimulating diverse learning outcomes. Whether such a change in the design of the program might enhance its impact would be a very worthwhile area to explore in future research.

¹ The CMIN-value is not taken into account since we had 714 observations for our CFA (in a CFA only the data from students who filled in all items of the questionnaire can be taken into account). Only when the number of observations is between 200 and 500 the CMIN-index can be used to examine the goodness-of-fit (Bollen, 1989; Hoyle, 1995).

² The SPQ is an earlier version of the R-SPQ-2F.

³ We considered the skewness and kurtosis of the dependent variables. Absolute values for skewness as well as for kurtosis were between .006 and 1.277.

⁴ We notice that from a teacher perspective the characteristics of students create a specific teaching context, and so are perceived as context characteristics.

⁵ This characteristic refers to the distinction between first year students and non first year students as well as to the differences in academic level between students, specified in terms of global end mark obtained the previous academic year.

References

- Baldwin, T. T., & Ford, K. J. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.
- Biggs, J., Kember, D., & Leung, D.Y.P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*, 133-149.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Brawner, C., Felder, R., Allen, R., & Brent, R. (2002). A survey of faculty teaching practices and involvement in faculty development activities. *Journal of Engineering Education, 91*, 393-396.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & R. Stine (Eds.), *Testing structural equation models* (pp.136-162). Newbury Park, CA: Sage.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of professional growth. *Teaching and Teacher Education, 18*, 947-967.
- Coffey, M., & Gibbs, G. (2000). Can academics benefit from training? Some preliminary evidence. *Teaching in Higher Education, 5*, 385-389.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Eison, J., & Stevens, E. (1995). Faculty development workshops and institutes. In W.A. Wright, & Associates, *Teaching improvement practices*. Boston, MA: Anker.
- Entwistle, N.J. (1998). Improving teaching through research on student learning. In J.J.F. Forest (Ed.) *University teaching: international perspectives* (pp.73-112). New York: Garland.
- Entwistle, N.J. (2005). Enhancing teaching-learning environments in undergraduate courses in electronic engineering: an introduction to the ETL project. *International Journal of Electrical Engineering Education, 42*, 1-7.
- Fishman, B.J., Marx, R.W., Best, S., & Tal, R.T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*, 643-658.
- Gibbs, G., & Coffey, M. (2000). Training to teach in higher education: A research agenda. *Teacher development, 4* (1), 31-44.
- Gibbs, G., & Coffey, M. (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education, 5*, 87-100.
- Gijbels, D., Van de Watering, G., Dochy, F., & Van den Bossche, P. (2005). The relationship between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education, 10*(4), 327-341.
- Gilbert, A., & Gibbs, G. (1999). A proposal for an international collaborative research programme to identify the impact of initial training on university teachers. *Research and Development in Higher Education, 21*, 131-143.
- Guskey, R.T. (2000). *Evaluating professional development*. Corwin Press: California.

- Ho, A. S. P. (1998). Changing teachers' conceptions of teaching as an approach to enhancing teaching and learning in tertiary education. *Unpublished PhD. Thesis*. Hong Kong: University of Hong Kong.
- Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education*, 42, 143-169.
- Hoyle, R. H. (Ed.). (1995). *Structural equation modelling: concepts, issues and applications*. Thousand Oakes: Sage.
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, New Jersey/London: Lawrence Earlbaum Associates Publishers.
- Joyce, B., & Showers, B. (1988). *Student achievement through staff development*. New York: Longman.
- Kember, D., & Leung, D.Y.P. (1998). The dimensionality of approaches to learning: An investigation with confirmatory factor analysis on the structure of the SPQ and LPQ. *British Journal of Educational Psychology*, 68, 395-407.
- Lea, S.J., Stephenson, D., & Troy, J. (2003). Higher education students' attitudes toward student-centred learning: Beyond 'educational bulimia'? *Studies in Higher Education*, 28(3), 321-334.
- Murray, K. & McDonald, R. (1997). The disjunction between lecturers' conceptions of teaching and their claimed educational practice. *Higher Education*, 33, 331-349.
- Norton, L., Richardson, J. T. E., Hartley, J., Newstead, S. & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher Education*, 50, 537-571.
- 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, 921-958.
- Postareff, L., Lindblom-Ylänne, S. & Nevgi, A. (2007). The effect of pedagogical training on teaching in higher education. *Teaching and Teacher Education*, 23, 557-571.
- Prosser, M., & Trigwell, K. (1997). Relations between perceptions of the teaching environment and approaches to teaching. *British Journal of Educational Psychology*, 67, 25-35.
- Ramsden, P. (1991). A performance indicator of teaching quality: The course experience questionnaire. *Higher Education*, 16, 129-150.
- Ramsden, P. (1997). The context of learning in academic departments. In F. Marton, D.J. Hounsell, and N.J. Entwistle (Eds.), *The experience of learning* (2nd ed.) (pp.198-216). Edinburgh: Scottish Academic Press.
- Richardson, J.T.E. (2006). Investigating the relationship between variations in students' perceptions of their academic environment and variations in study behaviour in distance education. *British Journal of Educational Psychology*, 76, 867-893.
- Rust, C. (1998). The impact of educational development workshops on teachers' practice. *International Journal for Academic Development*, 3 (1), 72-81.
- Shayer, M. (1992). Problems and issues in intervention studies. In: Demetriou, A., Shayer, M., and Efklides, A. (Eds.), *Neo-Piagetian theories of cognitive development: implications and applications for education* (pp.107-121). London: Routledge.

- Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational leadership*, 53, 12-16.
- Snelgrove, S., & Slater, J. (2003). Approaches to learning: Psychometric testing of a study process questionnaire. *Journal of Advanced Nursing*, 43(5), 496-505.
- Stes, A., Clement, M., & Van Petegem, P. (2007a). The effectiveness of a faculty training program: Long term and institutional impact. *International Journal for Academic Development*, 12, 99-109.
- Stes, A., Coertjens, L., & Van Petegem, P. (2007b). *Impact of a generic instructional development programme on teaching approach and behaviour*. Paper presented at the 12th Biennial Conference for Research on Learning and Instruction, Budapest.
- Thompson, D. E., Brooks, K., & Lizarraga, E. (2003). Perceived transfer of learning: From the distance education classroom to the workplace. *Assessment and Evaluation in Higher Education*, 28, 539-547.
- Trigwell, K., & Prosser, M. (1996). Congruence between intention and strategy in science teachers' approach to teaching. *Higher Education*, 32, 77-87.
- Trowler, P., & Cooper, A. (2002). Teaching and learning regimes: Implicit theories and recurrent practices in the enhancement of teaching and learning through educational development programmes. *Higher Education Research & Development*, 21, 221-240.
- Van Petegem, P., Stes, A., & Donche, V. (2007, August). *Teaching approaches in higher education: The influence of personal and contextual factors*. Poster presented at the 12th Biennial Conference for Research on Learning and Instruction, Budapest.
- Watkins, D., & Regmi, M. (1996). Towards the cross-cultural validation of a Western model of student approaches to learning. *Journal of Cross-Cultural Psychology*, 27, 547-560.
- Willcoxson, L. (1998). The impact of academics' learning and teaching preferences on their teaching practices: a pilot study. *Studies in Higher Education*, 23, 59-70.
- Wilson, S.M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In: A. Iron-Nejad, & P.D. Pearson (Eds.), *Review of Research in Education* (pp. 173-209). Washington, D.C.: AERA.
- Zhang, L.F. (2000). University students' learning approaches in three cultures: An investigation of Biggs 3P model. *Journal of Psychology*, 134, 37-56.

Table 1
Distribution of the experimental and control teachers according to the context characteristics, number of students in the classroom and students' expertise level

Context characteristic	Experimental teachers	Control teachers	Total group of participants
<i>Teaching a group of:</i>			
≤ 30 students	8	8	16
>30 but ≤ 100 students	7	7	14
> 100 students	4	3	7
	19	18	37
<i>Teaching:</i>			
First year students	6	7	13
Non First year students	13	11	24
	19	18	37

The distribution is shown for the pre-test.

There was a slight difference between the context pre (≤ 30 students) and post (>30 but ≤ 100 students) in the case of one teacher.

Table 2
Distribution of the experimental and control teachers according to the teacher characteristics of teaching discipline, gender, academic status and amount of teaching experience

Teacher characteristic	Experimental teachers	Control teachers	Total group of participants
<i>Belonging to a:</i>			
Hard discipline (e.g. chemistry or medicine)	14	11	25
Soft discipline (e.g. history or education)	5	7	12
	19	18	37
<i>Gender:</i>			
male	12	14	26
female	7	4	11
	19	18	37
<i>Academic status:</i>			
Assistant lecturer with PhD	8	3	11
(senior) lecturer or professor	11	13	24
unknown		2	2
	19	18	37
<i>Teaching experience:</i>			
≤ 5 years	9	5	14
> 5 years	10	12	22
unknown		1	1
	19	18	37

The distribution is shown for the pre-test.

Table 3
Scales of the three questionnaires used, listed with a typical item

Questionnaire	Scale	Typical item
ETLQ (part 2)	Aims and congruence	What was taught during classes seemed to match what we were supposed to learn.
	Teaching for understanding	The way this course was taught encouraged me to rethink my understanding of some aspects of the subject.
	Set work and feedback	The feedback given on my tasks and assignments helped to clarify what I hadn't fully understood.
	Assessing understanding	You had to really understand the subject to get good marks on this course.
	Staff enthusiasm and support	The teacher tried to share his enthusiasm for the subject with us.
	Student support	Talking with other students helped me to develop my understanding.
	Interest and enjoyment	Most of what I learned on this course was really interesting.
R-SPQ-2F	Studying with as little effort as possible	My aim for this course was to pass while doing as little work as possible.
	Learning by heart	I learned some things by rote even if I did not understand them.
	Studying is interesting	I found that studying for this course was at times as exciting as a good novel or movie.
	Spending extra time on studying	I found most new topics of this course interesting and often spent extra time trying to obtain more information about them.
ETLQ (part 4) + 8 items added by the researchers	Affective learning outcomes	I am better aware now of why I chose this degree course.
	Psychomotor learning outcomes	I now reflect on the way I study more often.
	Generic and information skills	I improved my ability to track down information in this subject area.
	Knowledge and subject-specific skills	I have obtained more understanding about the topics covered.

Table 4
 Mean scale scores and standard deviations for the experimental and control group at the pre-test

	Mean pre-test score for the experimental group	Mean pre-test score for the control group
<i>Scales concerning students' perceptions of teaching</i>		
Aims and congruence	.061 (.997)	-.041 (1.01)
Teaching for understanding	.015 (.996)	.057 (1.063)
Set work and feedback	.106 (1.027)	-.066 (1.036)
Assessing understanding	.053 (.998)	-.021 (0.990)
Staff enthusiasm and support	.035 (.974)	-.056 (1.097)
Student support	.049 (.998)	-.036 (1.013)
Interest and enjoyment	.056 (1.018)	.035 (1.028)
<i>Heading scales concerning students' study approach</i>		
Surface approach	-.032 (0.937)	.017 (1.028)
Deep approach	.108 (1.022)	-.019 (1.020)
<i>Sub-scales concerning students' study approach</i>		
Studying with minimal effort	-.035 (0.958)	-.012 (1.015)
Learning by heart	-.025 (0.961)	.070 (1.057)
Studying is interesting	.092 (1.034)	.020 (1.007)
Spending extra time	.083 (0.993)	-.016 (1.009)
<i>Scales concerning students' learning outcomes</i>		
Affective learning outcomes	.141 (0.970)	-.010 (1.049)
Psychomotor learning outcomes	.215 (1.010)	-.138 (0.984)
Generic and information skills	.141 (0.988)	-.094 (1.075)
Knowledge and subject-specific skills	.014 (1.076)	-.076 (1.048)

Standard deviations between brackets.

Table 5
Output data for the null model

	Variance at teacher level with regard to the post-test score	Variance at teacher level with regard to the difference score	Intra-class correlation (ICC)	Correlation between the variance at teacher level with regard to the post-test score and those with regard to the difference score
<i>Scales concerning students' perceptions of teaching</i>				
Aims and congruence	.212 (.062)	.105 (.044)	21%	.382*
Teaching for understanding	.177 (.053)	.121 (.048)	18%	.173
Set work and feedback	.348 (.100)	.165 (.067)	31%	.308*
Assessing understanding	.288 (.081)	.230 (.076)	29%	.478*
Staff enthusiasm and support	.210 (.060)	.105 (.043)	23%	.137
Student support	.213 (.062)	.036 (.025)	20%	.193
Interest and enjoyment	.278 (.076)	.089 (.038)	29%	.279
<i>Heading scales concerning students' study approach</i>				
Surface approach	.242 (.072)	.173 (.064)	24%	.605**
Deep approach	.257 (.074)	.143 (.056)	25%	.300*
<i>Sub-scales concerning students' study approach</i>				
Studying with minimal effort	.194 (.060)	.109 (.047)	19%	.562**
Learning by heart	.183 (.057)	.150 (.058)	19%	.515**
Studying is interesting	.230 (.067)	.133 (.053)	23%	.293
Spending extra time	.294 (.083)	.079 (.038)	26%	.449*
<i>Scales concerning students' learning outcomes</i>				
Affective learning outcomes	.147 (.047)	.061 (.033)	15%	-.070
Psychomotor learning outcomes	.206 (.059)	.015 (.018)	21%	-.419*
Generic and information skills	.247 (.070)	.065 (.034)	25%	-.129
Knowledge and subject-specific skills	.097 (.034)	.106 (.045)	11%	-.225

Significant effects are in bold.

* = moderate correlation; ** = strong correlation (Cohen, 1988)

Table 6
Output data for the gross and net models (effect of participation in training)

	The gross model		The net model	
	Participation in training (vs no training)	Participation in training * difference score	Participation in training (vs no training)	Participation in training * difference score
<i>Scales concerning students' perceptions of teaching</i>				
Aims and congruence	.023 (.168)	.068 (.149)	-.088 (.149)	-.079 (.152)
Teaching for understanding	-.141 (.156)	.042 (.153)	-.282 (.135)*	-.163 (.112)
Set work and feedback	.163 (.209)	.210 (.186)*	-.170 (.196)	-.111 (.159)
Assessing understanding	-.033 (.193)	.046 (.190)	-.190 (.191)	-.105 (.197)
Staff enthusiasm and support	.038 (.164)	.113 (.145)	-.090 (.148)	-.051 (.120)
Student support	-.017 (.167)	.014 (.116)	-.072 (.150)	-.050 (.116)
Interest and enjoyment ^o	-.221 (.182)*	-.033 (.137)	-.329 (.157)*	-.183 (.107)
<i>Heading scales concerning students' study approach</i>				
Surface approach	.041 (.183)	-.050 (.178)	.284 (.169)*	.182 (.194)
Deep approach	-.133 (.184)	-.073 (.165)	-.173 (.158)	-.079 (.150)
<i>Sub-scales concerning students' study approach</i>				
Studying with minimal effort	.080 (.167)	-.036 (.155)	.275 (.151)*	.178 (.166)
Learning by heart	-.006 (.164)	.009 (.169)	.217 (.156)*	.154 (.184)
Studying is interesting	-.242 (.173)*	-.118 (.157)	-.312 (.153)*	-.208 (.147)*
Spending extra time	-.086 (.193)	-.105 (.139)	-.027 (.163)	.044 (.117)
<i>Scales concerning students' learning outcomes</i>				
Affective learning outcomes	-.256 (.145)*	-.214 (.116)*	-.270 (.142)*	-.236 (.104)*
Psychomotor learning outcomes	.012 (.164)	-.071 (.100)	-.023 (.134)	-.082 (.118)
Generic and information skills	-.151 (.177)	-.133 (.127)	-.243 (.175)*	-.123 (.120)
Knowledge and subject-specific skills	-.159 (.128)	-.007 (.146)	-.262 (.130)*	-.111 (.121)

Significant effects are in bold.

* = small effect

^o The net model turned out to be too specific for the scale 'Interest and enjoyment'. 2 components which proved not to be significant in the net models for the other scales concerning students' perceptions of teaching were, therefore, excluded here (namely 1st year students * difference score, end mark obtained previous year * difference score).

Table 7
Output data for the interaction model relating to 1st years/no 1st years

	Experimental group teaching no 1 st years (vs control group teaching no 1 st years)	Experimental group teaching no 1 st years * difference score	Experimental group teaching 1 st years (vs control group teaching 1 st years)	Experimental group teaching 1 st years * difference score
<i>Scales concerning students' perceptions of teaching</i>				
Aims and congruence	-.163 (.183)	.014 (.192)	.261 (.300)*	.290 (.314)*
Teaching for understanding	-.292 (.166)*	-.077 (.146)	.065 (.282)	.222 (.234)*
Set work and feedback	-.062 (.231)	-.016 (.210)	-.301 (.389)*	.221 (.331)*
Assessing understanding	-.311 (.232)*	-.141 (.244)	.357 (.384)*	-.064 (.412)
Staff enthusiasm and support	-.241 (.180)*	-.022 (.154)	.481 (.305)*	.093 (.245)
Student support	.007 (.183)	-.030 (.151)	-.238 (.312)*	.042 (.240)
Interest and enjoyment ^o	-.463 (.180)*	-.236 (.119)*	.443 (.305)*	-.160 (.173)
<i>Heading scales concerning students' study approach</i>				
Surface approach	.437 (.202)*	.313 (.238)*	-.396 (.325)*	.325 (.396)*
Deep approach	-.303 (.186)*	-.261 (.167)*	.384 (.314)*	-.524 (.269)**
<i>Sub-scales concerning students' study approach</i>				
Studying with minimal effort	.407 (.181)*	.265 (.206)*	-.343 (.292)*	.204 (.336)*
Learning by heart	.364 (.186)*	.334 (.221)*	-.396 (.298)*	.492 (.367)*
Studying is interesting	-.463 (.181)*	-.320 (.177)*	.433 (.300)*	-.296 (.285)*
Spending extra time	-.053 (.199)	-.048 (.150)	.041 (.346)	-.237 (.238)*
<i>Scales concerning students' learning outcomes</i>				
Affective learning outcomes	-.315 (.176)*	-.210 (.156)*	.151 (.296)	.071 (.246)
Psychomotor learning outcomes	-.120 (.167)	-.128 (.153)	.278 (.279)*	-.105 (.240)
Generic and information skills	-.223 (.212)*	-.128 (.155)	-.068 (.371)	-.014 (.243)
Knowledge and subject-specific skills	-.374 (.162)*	-.268 (.156)*	.278 (.269)*	-.382 (.246)*

Significant effects are in bold.

* = small effect; ** = medium effect

^o Just as in the case of the net model, the interaction model turned out to be too specific for the scale 'Interest and enjoyment'. The same 2 components, which were excluded in the net model were, thus, also excluded here.

Table 8
Output data for the interaction model relating to the number of students

	Experimental group teaching a small number (vs control group teaching a small number)	Experimental group teaching a small number * difference score	Experimental group teaching a medium number (vs control group teaching a medium number)	Experimental group teaching a medium number * difference score	Experimental group teaching a large number (vs control group teaching a large number)	Experimental group teaching a large number * difference score
<i>Scales concerning students' perceptions of teaching</i>						
Aims and congruence	-.042 (.282)	.309 (.296)*	-.050 (.352)	.483 (.357)*	.051 (.400)	.483 (.357)*
Teaching for understanding	-.439 (.259)*	-.131 (.260)	.282 (.324)*	-.067 (.305)	.137 (.366)	.247 (.306)*
Set work and feedback	.329 (.331)*	.574 (.314)**	-.504 (.422)**	.921 (.372)***	-.955 (.480)***	.754 (.379)**
Assessing understanding	-.180 (.328)	-.082 (.358)	-.237 (.418)*	.141 (.450)	.588 (.490)**	-.210 (.513)*
Staff enthusiasm and support	-.079 (.279)	.005 (.265)	-.072 (.353)	-.022 (.312)	.203 (.410)*	.267 (.317)*
Student support	-.464 (.289)*	-.471 (.269)*	.457 (.364)*	-.360 (.313)*	.524 (.422)**	-.740 (.317)**
Interest and enjoyment ^o	-.503 (.286)**	-.104 (.260)	.248 (.365)*	.029 (.306)	.490 (.411)*	.249 (.292)*
<i>Heading scales concerning students' study approach</i>						
Surface approach	.431 (.306)*	.134 (.362)	-.089 (.380)	-.157 (.450)	-.453 (.430)*	.108 (.512)
Deep approach	-.572 (.282)**	-.426 (.298)*	.468 (.352)*	-.359 (.358)*	.665 (.398)**	-.579 (.377)**
<i>Sub-scales concerning students' study approach</i>						
Studying with minimal effort	.507 (.280)**	.230 (.326)*	-.203 (.343)*	-.009 (.396)	-.542 (.378)**	-.009 (.396)
Learning by heart	.289 (.288)*	.091 (.346)	-.003 (.356)	-.200 (.428)*	-.343 (.401)*	.193 (.481)
Studying is interesting	-.689 (.279)**	-.416 (.299)*	.467 (.346)*	-.178 (.358)	.613 (.389)**	-.436 (.378)*
Spending extra time	-.334 (.303)*	-.433 (.269)*	.342 (.390)*	-.568 (.315)**	.421 (.462)*	-.570 (.315)**
<i>Scales concerning students' learning outcomes</i>						
Affective learning outcomes	-.522 (.280)**	-.156 (.284)	.349 (.343)*	.108 (.330)	.429 (.381)*	.111 (.329)
Psychomotor learning outcomes	.214 (.267)*	-.144 (.278)	.074 (.324)	-.020 (.322)	.730 (.353)**	-.144 (.322)
Generic and information skills	-.486 (.323)*	-.099 (.283)	.365 (.412)*	.158 (.328)	.350 (.486)*	-.195 (.326)
Knowledge and subject-specific skills	-.489 (.259)*	.094 (.282)	.375 (.313)*	.326 (.327)*	.227 (.341)*	.119 (.328)

Small number: ≤30 students; Medium number: >30 but ≤100 students; Large number: >100 students

Significant effects are in bold.

* = small effect; ** = medium effect; *** = large effect

^o Just as in the case of the net model and with the first interaction model, this model turned out to be too specific for the scale 'Interest and enjoyment'. The same 2 components, which were excluded in the net model and the first interaction model were, thus, also excluded here. Since this resulted in a model that still proved to be too specific, the components 'soft discipline' and 'soft discipline * difference score' were also excluded.

Appendix 1a
Output data for the net model relating to teacher characteristics

	Teacher characteristic									
	Soft discipline (vs hard)	Soft discipline * difference score	Female gender (vs male)	Female gender * difference score	Status of assistant with PhD (vs lecturer/ professor)	Status of assistant with PhD * difference score	>5 years of experience (vs ≤5 years)	>5 years of experience * difference score	Participation in training (vs no training)	Participation in training * difference score
<i>Scales concerning students' perceptions of teaching</i>										
Aims and congruence	-.022 (.174)	-.022 (.175)	.168 (.162)	.153 (.162)	.091 (.158)	-.111 (.165)	-.544 (.143)	-.400 (.143)	-.088 (.149)	-.079 (.152)
Teaching for understanding	-.216 (.158)	.107 (.128)	-.035 (.143)	.081 (.108)	.021 (.147)	-.002 (.123)	-.385 (.129)	-.500 (.098)	-.282 (.135)	-.163 (.112)
Set work and feedback	-.393 (.222)	.125 (.176)	.394 (.205)	.447 (.158)	.345 (.207)	.114 (.167)	-.541 (.184)	-.774 (.141)	-.170 (.196)	-.111 (.159)
Assessing understanding	-.185 (.222)	.301 (.226)	.347 (.211)	.291 (.216)	.070 (.200)	.363 (.211)	-.339 (.184)	-.320 (.189)	-.190 (.191)	-.105 (.197)
Staff enthusiasm and support	-.239 (.173)	.031 (.138)	.122 (.160)	.199 (.120)	.180 (.160)	-.051 (.130)	-.353 (.142)	-.615 (.107)	-.090 (.148)	-.051 (.120)
Student support	-.286 (.176)	.072 (.132)	.049 (.162)	.316 (.111)	.394 (.163)	.299 (.125)	.119 (.143)	-.133 (.100)	-.072 (.150)	-.050 (.116)
Interest and enjoyment	-.088 (.177)	.232 (.116)	-.110 (.171)	-.018 (.107)	.199 (.171)	-.069 (.117)	-.328 (.153)	-.480 (.095)	-.329 (.157)	-.183 (.107)
<i>Heading scales concerning students' study approach</i>										
Surface approach	.310 (.196)	.127 (.223)	-.167 (.185)	-.370 (.212)	-.090 (.175)	-.005 (.207)	.220 (.161)	.248 (.186)	.284 (.169)	.182 (.194)
Deep approach	-.239 (.185)	.348 (.173)	-.097 (.171)	.153 (.158)	-.013 (.168)	-.040 (.162)	-.216 (.152)	-.338 (.140)	-.173 (.158)	-.079 (.150)
<i>Sub-scales concerning students' study approach</i>										
Studying with minimal effort	.289 (.177)	.074 (.191)	-.125 (.164)	-.335 (.178)	-.068 (.158)	-.044 (.178)	.162 (.144)	.187 (.157)	.275 (.151)	.178 (.166)
Learning by heart	.141 (.181)	.053 (.210)	-.215 (.170)	-.260 (.199)	-.209 (.162)	.098 (.197)	.191 (.148)	.251 (.175)	.217 (.156)	.154 (.184)
Studying is interesting	-.160 (.178)	.302 (.169)	-.035 (.165)	.176 (.154)	.117 (.162)	-.016 (.158)	-.173 (.146)	-.348 (.137)	-.312 (.153)	-.208 (.147)
Spending extra time	-.135 (.190)	.278 (.133)	-.096 (.174)	.061 (.111)	-.142 (.177)	-.196 (.126)	-.187 (.156)	-.268 (.102)	-.027 (.163)	.044 (.117)
<i>Scales concerning students' learning outcomes</i>										
Affective learning outcomes	.009 (.167)	.164 (.137)	-.232 (.150)	.026 (.115)	.044 (.154)	-.212 (.131)	-.097 (.135)	-.263 (.104)	-.270 (.142)	-.236 (.104)
Psychomotor learning outcomes	-.294 (.158)	.025 (.134)	.110 (.141)	.115 (.112)	.071 (.146)	-.093 (.128)	.063 (.127)	-.159 (.102)	-.023 (.134)	-.082 (.118)
Generic and information skills	-.295 (.204)	.104 (.136)	.001 (.188)	.313 (.114)	.186 (.190)	-.131 (.129)	-.093 (.168)	-.217 (.103)	-.243 (.175)	-.123 (.120)
Knowledge and subject-specific skills	-.161 (.153)	.168 (.137)	-.225 (.137)	-.014 (.115)	.184 (.141)	-.181 (.131)	-.276 (.123)	-.419 (.105)	-.262 (.130)	-.111 (.121)

Although the output data relating to teacher characteristics on the one hand and those relating to context and student characteristics on the other are shown in separate appendices, they are all part of one and the same 'net model'.
Significant effects are in bold.

Appendix 1b

Output data for the net model relating to context and student characteristics

	Context characteristic						Student characteristic		
	Medium number of students (vs low number)	Medium number of students *	Large number of students (vs low number)	Large number of students *	1 st year students (vs no 1 st year students)	1 st year students * difference score	Female gender (vs male)	Age (in years)	End mark obtained year before (in %)
<i>Scales concerning students' perceptions of teaching</i>									
Aims and congruence	.021 (.171)	.207 (.180)	-.256 (.244)	.452 (.246)	.051 (.181)	.091 (.158)	.095 (.046)	-.004 (.009)	-.001 (.001)
Teaching for understanding	-.312 (.158)	.091 (.146)	.011 (.227)	.340 (.186)	-.407 (.188)	.040 (.134)	-.032 (.048)	.026 (.009)	-.001 (.001)
Set work and feedback	-.275 (.221)	.355 (.197)	.068 (.320)	.573 (.246)	-.425 (.261)	-.095 (.186)	-.027 (.052)	.011 (.010)	-.002 (.002)
Assessing understanding	-.247 (.214)	.047 (.224)	-.202 (.315)	.189 (.320)	.137 (.246)	-.035 (.237)	.140 (.045)	.023 (.009)	-.001 (.001)
Staff enthusiasm and support	-.337 (.171)	.256 (.151)	-.072 (.249)	.340 (.195)	-.388 (.201)	.083 (.141)	.030 (.046)	.002 (.009)	.000 (.001)
Student support	-.177 (.175)	.356 (.151)	.253 (.253)	.631 (.189)	-.112 (.205)	-.054 (.135)	.090 (.049)	-.021 (.009)	.001 (.001)
Interest and enjoyment	-.059 (.179)	.202 (.140)	-.364 (.247)	.264 (.142)	-.341 (.155)	°	.044 (.044)	.007 (.009)	°
<i>Heading scales concerning students' study approach</i>									
Surface approach	.031 (.191)	-.333 (.222)	.002 (.274)	-.352 (.316)	.151 (.222)	-.142 (.232)	-.388 (.048)	-.015 (.010)	.000 (.001)
Deep approach	-.321 (.182)	-.103 (.179)	-.472 (.263)	-.372 (.243)	-.107 (.216)	.283 (.179)	-.068 (.049)	.067 (.010)	-.001 (.002)
<i>Sub-scales concerning students' study approach</i>									
Studying with minimal effort	-.037 (.174)	-.315 (.194)	-.037 (.246)	-.365 (.268)	.181 (.205)	-.081 (.197)	-.391 (.050)	-.020 (.010)	.001 (.002)
Learning by heart	.195 (.177)	-.223 (.211)	.227 (.252)	-.207 (.299)	.018 (.208)	-.221 (.220)	-.265 (.048)	-.005 (.010)	-.001 (.001)
Studying is interesting	-.138 (.175)	.059 (.176)	-.445 (.252)	-.112 (.237)	-.149 (.208)	.220 (.174)	.057 (.049)	.052 (.010)	-.002 (.001)
Spending extra time	-.279 (.187)	-.206 (.150)	-.421 (.279)	-.520 (.193)	-.098 (.227)	.312 (.139)	-.195 (.050)	.068 (.010)	-.001 (.002)
<i>Scales concerning students' learning outcomes</i>									
Affective learning outcomes	-.159 (.169)	.083 (.158)	-.451 (.241)	.094 (.200)	-.111 (.203)	.004 (.143)	.034 (.051)	.026 (.011)	-.002 (.002)
Psychomotor learning outcomes	-.193 (.161)	.017 (.154)	-.191 (.227)	.169 (.196)	.555 (.193)	-.081 (.140)	.038 (.050)	.033 (.010)	-.003 (.002)
Generic and information skills	-.267 (.202)	.203 (.157)	-.033 (.302)	.163 (.199)	.020 (.244)	.130 (.142)	.024 (.051)	.027 (.011)	-.003 (.002)
Knowledge and subject-specific skills	-.159 (.156)	.111 (.157)	-.157 (.219)	.219 (.201)	-.035 (.190)	.215 (.143)	.070 (.051)	.021 (.010)	-.001 (.002)

Although the output data relating to teacher characteristics on the one hand and those relating to context and student characteristics on the other are shown in separate appendices, they are all part of one and the same 'net model'.

Significant effects are in bold.

° The net model turned out to be too specific for the scale 'Interest and enjoyment'. 2 components which proved not to be significant in the net models for the other scales concerning students' perceptions of teaching were excluded here (namely 1st year students * difference score, end mark obtained previous year * difference score).