Improving university lectures with feedback and consultation
Knol, M.H.

Citation for published version (APA):

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Introduction
Faculty development is hot, research on faculty development is not

Accountability at universities, towards the end of promoting excellence in research and education, has become more and more important over the years. Enders, De Boer and Weyer (2012) provided an overview of the current situation of universities in the Netherlands. They state that at present roughly two-third of the budget of Dutch universities derives from the government. Universities are seriously dependent on this income stream that comes with strings attached as regards funding by student numbers. Dutch universities are required to have an internal and external evaluation system for both teaching and research. In addition, other internal and external accountability requirements, e.g., quality assurance schemes, accreditation, and performance monitoring, have been introduced over the years (Enders, De Boer & Weyer, 2012).

To excel in research, university professors (meaning part-time and full-time assistant, associate and full professors) start with extensive training through PhD projects (in the Netherlands, a PhD position is considered a proper salaried job). At the same time, the educational training of university professors is extremely limited. This state of affairs is similar to that in other countries. In an opinion piece in the International Journal for Academic Development, Baume (2006) stated that teaching in higher education may be one of the last non-professions. Unlike their colleagues in elementary and secondary education, most professors have little formal training in teaching. Mostly, they rely on their own past experience as teachers, and on the examples set by their professors when they themselves were students. University teaching therefore resembles on-the-job training, which often takes place in isolation, with little help, and no resources (Knapper & Piccinin, 1999). In the Netherlands, teachers are required to follow four years of teacher education to be able to teach in primary or secondary education. Master graduates are able to teach at secondary education after one additional year of teacher training. In contrast, to be able to teach at universities, one basically needs only to be a research expert on the subject matter.

At all educational levels, educational researchers agree that teaching not only requires subject matter knowledge, but also pedagogical content knowledge (knowledge on how to teach the subject matter, see Shulman, 1986) and teaching skills that distinguish teachers from subject matter specialists. Ramsden (2002) defined the activity of teaching in higher education to include the objectives of the curriculum, the pedagogies of conveying the knowledge that these objectives embody, the assessment of students, and the evaluation of the effectiveness of the teaching.

Thus, to be a subject matter specialist does not guarantee good teaching. Of course, some subject matter specialists at universities do excel in the quality of their teaching. In an
extensive meta-analysis based on 58 articles, Hattie and Marsh (1996) showed no correlation between the quality of research and the quality of teaching of university professors. Regardless of discipline, professors come in all different flavors; good at teaching and bad at research, good at teaching and good at research, bad at teaching and good at research, and bad at both. Related to this, Handal (1999) speaks of “dual professionalism”, and Baume (2006) noted: “I hope that, soon, teaching in higher education will be recognized for its difficulty, its importance and the great extent of its responsibility” (p. 58).

In response to lack in formal preparation for teaching, faculty development centers have been created, starting in the 1970s, to support and improve university teaching. In addition, the presidents of all Dutch universities signed an agreement in 2008 on the instatement of a basic quality of teaching certificate for university professors. The certificate concerns new standards of teaching, and is recognized at all fourteen Dutch universities (VSNU, 2008). In other Western countries, similar initiatives have been taken, e.g., the UK introduced a teaching qualification which meets new standards of teaching, and needs to be obtained by all new teaching staff from 2006 (DfES, 2003, cited in Baume, 2006).

In addition to these developments, studies on effective interventions to improve the quality of university teaching have become equally important. Unfortunately to date, research funding on (effective interventions in) higher education has been limited. For example, the Dutch Programme Council for Educational Research (PROO) funds educational research on primary education, general secondary education, (pre-) vocational education and teacher training institutes (training primary and secondary education teachers). However, research on higher education is excluded (PROO, 2012). As a result, the effectiveness of faculty development practices is seldom investigated thoroughly. In other western countries, research grants have been allocated to this matter, but, compared to other educational fields, this field of research is largely neglected. In each review of the literature on the effectiveness of faculty development practices, authors stressed the importance of more research, and specifically more experimental research in this field (Levinson-Rose & Menges, 1981; Prebble et al., 2004; Steinert et al., 2006; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010; Weimer & Lenze, 1997). Each of these reviews concluded that many studies in this field are characterized by important limitations. Some of the most important limitations are the following.

First, studies are often limited to small and/or selected samples. For example, participants are often professors who approached faculty development centers with the aim to improve their teaching, and were therefore often well motivated to change from the start. In most cases, results show positive effects, but whether the effects are due to the intervention, the initial motivation, or simply time on task remains unknown. Second, the evaluations
are generally limited to measures of satisfaction of the participants. Third, participants are often not randomly assigned to various interventions, leaving the specific and relative effects unclear. Fourth, effect studies often lack a control condition, leaving the relative effects compared to no intervention unclear. Fifth, studies often lack thorough research on the psychometric quality of the instruments used to evaluate improvements.

In summary, with the increasing acknowledgement of (and investment in) university teaching, additional experimental research on faculty development practices on multiple levels of evaluation has become indispensable. Below, the various levels of evaluation are discussed.

**Considering levels of evaluation**

Kirkpatrick (1994) distinguished four levels of evaluation of training programs in business and industry: reaction, learning, behavior, and results. The reaction level of evaluation concerns participants’ satisfaction with the program. The learning level concerns the knowledge, attitudes, and skills that participants acquire as a result of the program. The behavioral level concerns participants’ behavioral changes on the job, due to the program, and the result level concerns the effects of the program on the organization. Guskey (2000) adapted Kirkpatrick’s evaluation model to the educational field, specifically to evaluate the professional development of teachers. Guskey’s five-level evaluation model comprises participants’ reactions (level 1), participant’s learning (level 2), organizational support and change (level 3), participant’s application of new knowledge and skills (level 4), and student learning outcomes (level 5). Both models imply a hierarchic arrangement of levels, from simple to more complex, whereas each higher level builds on the preceding levels (Guskey, 2000). One example of more extensive quasi-experimental research, containing pre- and post-tests, control group comparison, and multiple levels of evaluation, are the studies of Stes and colleagues (2010, 2011). They found significant effects of their instructional development program on the teachers’ approaches to teaching (which relates to level 2 and 4) (Stes, Coertjens, & Van Petegem, 2010), but limited effects on students’ approaches to their study (which relates to an alternative result on the student outcome level 5) (Stes, De Maeyer, Gijbels, & Van Petegem, 2011). Such studies stress the importance of evaluation on multiple levels of effects. With more knowledge on the actual impact on various levels, faculty development centers can target and combine the optimal interventions, corresponding to prior aims for improvement.
The present aims

The first aim in the present dissertation is to overcome the limitations in previous research on faculty development interventions, as stated above, in an investigation of the effects of two specific faculty development interventions on university professors’ lectures (class meetings in which lecturing is the teaching format). Although lecturing is not the most popular teaching format used in education, I chose to focus on this format since it still constitutes a substantial, and often indispensible, part of regular teaching practices at universities (Lammers & Murphy, 2002). In addition, the format of lecturing is largely comparable over different departments, and so provides me with the opportunity to gather a substantial amount of data to investigate the effects of these interventions on professors’ teaching behavior and students’ self-reported learning.

The dissertation includes two experimental studies; a pilot study with twenty-five participants from a single department at the University of Amsterdam, and a larger study with seventy-five participants from a wide variety of departments at the same university. The participants were professors, who had not approached a faculty development center for support. Both studies included a control condition. In each experiment, professors were randomly assigned to the experimental and control conditions.

The second aim is to investigate the impact of the two interventions on Guskey’s levels one, two, four, and five, that is, the effects on professors’ self-reported satisfaction with the interventions (level 1), professors’ self reported learning (level 2), professors’ use of new knowledge and skills, measured by students’ evaluations of lecturing (level 4), and students’ self-assessed learning outcomes (level 5).

Below I present the rationale concerning the two faculty development interventions considered here and the main research question investigated in this dissertation. Next, I provide an outline of the dissertation. I end this dissertation with an overview and a discussion of the results.

Investigating student feedback and consultation

In the present research project, the objective was to investigate thoroughly the effects of student feedback provided to professors on their lectures, with and without additional individual consultations with professors. Aside from formal training programs and workshops, individual peer or expert consultation is one of the most commonly used interventions in faculty development, specifically instructional development with respect to small and larger classroom teaching (Knapper & Piccinin, 1999; Penny & Coe, 2004; Prebble et al., 2004). Based
on the available research, Lenze (1996) identified consultation as an instructional development strategy preferable to other approaches, such as workshops, grants for instructional improvement, advice from colleagues, and provision of resource materials. A common consultation procedure is for the consultant to clarify teaching goals, to encourage reflection about aims and methods, combined with some sort of feedback, to facilitate discussion on improvement strategies, and sometimes to conduct follow-up evaluation (Knapper & Piccinin, 1999; Penny & Coe, 2004). Feedback is gathered either through students’ evaluations of teaching (e.g., Rindermann, Kohler & Meisenberg, 2007), or by more extensive means, such as classroom observations (e.g., Wilson, 1986; Piccinin, Cristy & McCoy, 1999), videotaping (e.g., Rozeman & Kerwin, 1991), or arranged student focus groups (e.g., Piccinin, Cristy & McCoy, 1999; Coffman, 1998). In this dissertation, I focus on combining students’ evaluations of teaching (SETs) with individual consultation (SET consultation).

SETs are considered a potentially useful source of feedback (Prebble et al., 2004), and are valued as a formative feedback instrument by faculty members and faculty developers (Baxter, 1991; Schmelkin, Spencer & Gellman, 1997; Penny & Coe, 2004). Nowadays, collecting SETs at the end of the term or course has become common practice at universities worldwide. Unfortunately, despite the effort and despite its main purpose to provide faculty with feedback, collecting student feedback at the end of the term or course per se has little to no effect on teaching behavior (Hendry & Dean, 2002; Kember, Leung & Kwan, 2002; Marsh, 2007a). Providing professors with intermediate student feedback has some effect, in terms of an increase in SET ratings (Cohen, 1980; Menges & Brinko, 1989). Subsequently, augmenting SETs with individual consultation (SET consultation) has proven to be considerably more effective in various studies and meta-analyses (e.g. Cohen, 1980; Hampton & Reiser, 2004; Menges & Brinko, 1986; Marsh & Roche, 1993; Penny & Coe, 2004; Piccinin, Cristi & McCoy, 1999; Rindermann, Kohler & Meisenberg, 2007; Dresel & Rindermann, 2011). Consultation is considered an improvement to end of the term evaluations, since the latter often come too late to be of use, and generally come without any practical suggestions or support for change and improvement (Knapper & Piccinin, 1999; Penny & Coe, 2004).

I stated earlier that reviewers call for more in-depth research on the effects of faculty development interventions in general. In specific reviews on the effects of the two interventions investigated, intermediate SETs and SET consultation, this is also the case. One important finding, noted by these reviewers, is that the variation in effects of SET consultation is large (Menges & Brinko, 1986; Penny & Coe, 2004). The next step in this field of research is therefore to provide more insight into the effectiveness of particular approaches and procedures. In a meta-analysis, Penny and Coe (2004) studied the predictors of successful SET consultation, but
were limited in their research due to the small number of experimental studies. They noted: “Thus, the most robust finding may be that more research is needed” (p.236). In addition, they called for more research on SET consultation in settings other than North America, and replication of studies of various approaches to consultation. Furthermore, l’Hommedieu, Menges and Brinko (1990) provided a critical assessment of the limited effects of student feedback only. They focused on important methodological issues in previous research, and stated that the literature is hampered by pervasive threats to the internal and external validity of research findings. They stressed the need for further research to consolidate these findings. Some of their recommendations concern more adequate research on the instruments used, larger samples, sampling across subject areas and teacher characteristics, pre-tests, systematically assigned subjects and/or statistically controlling for moderating variables, studies of large lecture classes, consideration for the appropriate unit of analysis, and use of comparable measures (mid-term evaluation does not necessarily compare to end-of-the-term evaluation). In this dissertation these recommendations are taken into account.

The interventions investigated in this dissertation involve providing professors with SETs on their lectures with or without individual consultation during the course they are teaching. In the first experimental (pilot) study the aim was to investigate the effects of intermediate SETs with consultation. In the second experimental study the aim was to separate the effects of feedback and consultation by investigating intermediate SETs with and without consultation.

The main research question addressed in this dissertation is:

What are the effects of intermediate student feedback with and without consultation on professors’ self-reported satisfaction with the interventions (level 1), professors’ self reported learning on lecturing (level 2), professors’ lecturing skills, measured by students’ evaluations of lecturing (level 4), and students’ self-assessed learning outcomes (level 5)?

Based on the previous literature, I hypothesized that the effects of intermediate feedback on these levels of evaluation were small and the effects of intermediate feedback with consultation on these levels were medium to large.

The effects on student ratings data were investigated with multilevel regression analysis. This statistical approach accounts for the clustering in the data due to systematic differences between the lectures, the students, and the professors. The aim was to complement previous findings with new analyses, using this modern statistical approach.

Additionally, the effects on the teaching dimensions that were targeted for improvement during consultation were separated from the effects on non-targeted dimensions, to indicate whether the effects were due to the selected consultation approach or due to a Hawthorne
effect (i.e. due to the attention/social treatment one receives). Finally, the moderating effects of specific professor and course characteristics (i.e., professors’ age, professors’ prior quality of teaching and class size) were investigated. In doing so, I hope to contribute to the current body of knowledge in this field.

**Outline of this dissertation**

The chapters in this dissertation consist of accepted or submitted journal articles. Chapter 2 concerns an investigation on the psychometric quality of the instrument used to evaluate lectures, as assessed by students (the instrument measures various dimensions of lecturing skills), by means of confirmatory multilevel factor analysis. The chapter contains analyses on the construct validity, internal structure, and reliability of these teaching dimensions. These analyses are based on data collected in the second experiment of this dissertation. Furthermore, this chapter provides a theoretical framework on the relationship between the professors’ lecturing behavior and the students’ learning process. Test of these relationships are included in this chapter.

Chapter 3 provides a theoretical framework on the approach to consultation used in this study (collaborative consultation) based on theories on behavioral change. In addition, chapter 2 concerns a first experimental (pilot) study with twenty-five psychology professors, who were randomly assigned to either the experimental condition with SET consultation or the control condition with neither feedback nor consultation. The effects are studied in terms of changes in the professors SET results during the course (Guskey’s level 4: changes in behavior, according to students). Chapter 2 (investigating the instrument) and three (piloting the procedure and approach to intermediate SET consultation) are considered to be a preparation for the investigations conducted in chapter 4 and 5.

Chapter 4 concerns a second experiment with seventy-five professors from a wide variety of departments, who were randomly assigned to one of three conditions, a feedback-only condition, a feedback-plus-consultation condition, or a control condition. This chapter concerns the effects of the second experiment on Guskey’s level of evaluation 1 and 2; the effects on professors’ self-reported satisfaction with the interventions, and professors’ self reported learning.

Chapter 5 concerns the effects of the second experiment in terms of changes in the professors’ SET results during the course (Guskey’s level 4), and the effects in terms of changes in students’ self-assessed learning outcomes (Guskey’s level 5). The dissertation closes with a final discussion on the effectiveness of the chosen approach to intermediate feedback and intermediate feedback plus consultation.
References


