

Jean-Luc PATRY, Angela GASTAGER

Zalcburgo universitetas • University of Salzburg

KOKYBĖS VERTINIMO DILEMOS¹

DILEMMAS IN QUALITY ASSESSMENT¹

SANTRAUKA

Straipsnyje analizuojama potencialių vertybių ir konfliktų įvertinimo kriterijų struktūrinė klasifikacija. Tyrimo pagrindą sudaro konfliktų tipų išskyrimas (tikslų ir priemonių konfliktai; kokybiniai ir kiekybiniai konfliktai). Aptariamos šešios vertybių sritys: etinės, metodinės, socialinės ir interaktyviosios, taip pat įstatyminės, ekonominės ir asmeninės. Konfliktų, atsirandančių tarp šių vertybių struktūrinė klasifikacija taikyta konkrečiai įvertinant dėstymą, kurį atliko Zalcburgo universiteto studentai. Aptariami kai kurie konfliktai, pvz., aprėptis prieš patikimumą; individualios prieš socialines ir į kriterijus orientuotas standartines normas; studentų reitingavimas prieš įvertinimą, dėstytojų įtaka tyrimų rezultatams. Struktūrinę klasifikaciją siūloma taikyti prieš atliekant įvertinimą.

PAGRINDINIŲ SĄVOKŲ APIBRĖŽIMAI

- *Antagonizmas* – prieštarų principu, jėgų ar veiksmų santykis; šiuo atveju prieštarų tikslai, nesuderinamos priemonės.
- *Aprėptis* – šiuo atveju sričių, vertinamų taikant duotąsias priemones, skaičius.
- *Normos, orientuotos į kriterijus* – tokios normos, kai įvertinant studentus remiamasi iš anksto nustatytais kriterijais, t. y., atsižvelgiama į tikslų pasiekimo laipsnį.
- *Sprendimų priėmimas sudėtingomis socialinėmis aplinkybėmis* – diskusijos ir pasikeitimas argumentais už ir/ar prieš nesutampančias vertybes atskirų asmenų sąveikos ir komunikacijos procese.
- *Sprendimo priėmimo aplinkybės* – situacija, kurioje protagonistas prašomas priimti sprendimą dėl vieno ar daugiau veiksmų.
- *Dilema* – tokia sprendimo priėmimo situacija, kurioje bet koks priimtas sprendimas pažeis kokias nors kitas vertybes; paprastai tai kokybinė konfliktinė situacija.
- *Įvertinimo dilemos* – sudėtingos sprendimo priėmimo aplinkybės įvertinimo procese visuose švietimo sistemos sektoriuose.
- *Įvertinimo procesas (suaugusiųjų) švietime* – sistemingas faktų ir kitos informacijos rinkimas bei interpretavimas siekiant priimti vertybinį sprendimą apie veiklos pobūdį visuose (suaugusiųjų) švietimo srityse.
- *Sąžiningumas* – šiuo atveju protagonistas turi progą laikytis įvertinimo kriterijų reikalavimų (t. y. laiku sužino tuos kriterijus ir turi galimybę jų siekti, pvz., jam dėstomas tam tikras kursas, kurio įsisavinimas vėliau bus įvertintas).

ABSTRACT

A taxonomy of potential values conflicts in evaluations is presented. Basis is the distinction of types of conflict (aims vs. means conflicts; qualitative vs. quantitative conflicts). Six areas of values are discussed: the ethical, methodical, social and interactive, legal, economic, and personal values. The taxonomy of conflicts between these values is applied to a concrete evaluation process, namely the student rating of teaching at the University of Salzburg. Some of the conflicts are discussed, e.g. bandwidth vs. fidelity, individual vs. social and criterion oriented reference norm, student ratings vs. evaluation, and cheating teachers. It is suggested to use the taxonomy before starting an evaluation.

DEFINITIONS OF KEY WORDS

- *Antagonism*: relation between opposing principles or forces or factor; in this particular case: opposed goals, incompatible means.
- *Bandwidth*: here number of domains that are assessed with a given instrument.
- *Criterion oriented norm*: reference norm to evaluate students based on pre-determined criteria, i.e., referring to the degree of accomplishment of goals.
- *Decision-making in difficult social-situations*: discussion and interchange of arguments concerning the pro and/ or contra for values, which conflict in the interaction and communication of persons.
- *Decision situation*: situation in which the protagonist is asked to take a decision between two or more actions.
- *Dilemma*: decision situation in which any decision will violate some values; usually a qualitative conflict situation.
- *Dilemmas in evaluations*: (difficult) decision situations in evaluation processes of all possible sections concerning education.
- *Evaluation-process in (adult) education*: systematic collection and interpretation of facts and hints with the goal to achieve a decision of values about actions in all fields of (adult) education.
- *Fairness*: here the protagonist gets the opportunity to fulfill the requirements with respect to the criteria which are used for the evaluation (i.e., knows the criteria in time and has the possibility to achieve them, e.g., is taught the content which is evaluated later on).

¹ Iš dalies remiamasi projektu „Įvertinimas švietime“, grantas #P13594-SOZ, gautas iš Austrijos mokslo fondo (FWF).

¹ Partly based on the project “Evaluation in Education”, grant # P13594-SOZ by the Austrian Science Fund (FWF).

- *Patikimumas* – šiuo atveju tikslumas, vertinant savybę, taikant duotąją priemonę.
- *Formuojamasis įvertinimas* – įvertinimas pasitelkiamas siekiant daryti įtaką analizuojamam procesui.
- *Įvertinimo / švietimo tikslai* – jie apibūdina tai, ko kas nors siekia duotomis aplinkybėmis.
- *Priemonės įvertinime / švietime* – apibūdina, ką galima daryti norint pasiekti tikslą duotomis aplinkybėmis.
- *Kokybinė konfliktinė situacija* – tokia sprendimo priėmimo situacija, kai galimi veiksmai ar vertybių pasiekimas yra nesuderinami (*arba–arba*).
- *Kiekybinė konfliktinė situacija* – tokia sprendimo priėmimo situacija, kai galimi veiksmai gali būti apibūdinami sąvokomis *daučiau ar mažiau*; kuo labiau laikomasi vienos vertybės, tuo mažiau lieka vietos kitai.
- *Atsakomybė* – (1) jėga, verčianti jus laikytis įsipareigojimų; (2) veiksmų eiga, kurios ta jėga reikalauja, jai esant išorinei (pvz., kitų žmonių nustatytos normų sistemos) ar vidinei (pvz., protagonisto sąžine paremtos normų sistemos).
- *Atliepiamas įvertinimas (modelis)* – svarbios atliepiamojo įvertinimo proceso vertybės yra (1) skaidrumas apie visus įvertinimo žingsnius, (2) visų į vertinimo procesą įtrauktų asmenų dalyvavimas, pvz., dalinai priimant svarbius įvertinimo proceso sprendimus ir (3) tolesnės įtakos svarstymas, pvz., žinios apie įvertinimo kriterijus.
- *Socialiai reglamentuotos normos* – vertinimo norma, taikoma studentų įvertinimui, kai remiamasi socialiniais palyginimais, t. y. atsižvelgiama į jų darbo kokybę lyginant su kitų tos grupės darbo kokybe.
- *Sumuojamasis įvertinimas* – situacijos įvertinimas duotuoju momentu (ypač proceso pabaigoje).
- *Skaidrumas* – šiuo atveju svarbūs protagonistai, sukaupę visas žinias apie tikslus ir procesus.
- *Vertybės* – tai, kas kiekvienam svarbu; (teigiama ar neigiama) kokybė, tą svarbumą paverčia norimu ar vertinamu dalyku.
- *Vertybių konfliktas* – atskiras žmogus tuo pat metu turi savo asmeninių vertybių sistemą (vidinės vertybės), o skirtingi žmonės gali turėti skirtingas vertybes (tarpasmenines vertybes). Skirtingos vertybės gali būti konfliktu, kylančių įvairiomis veiklos aplinkybėmis, priežastis.

ĮVADAS

Žmonės, dirbantys dviejų ar kelių socialinių ryšių sandūrose, susiduria su daugybe dažnai vienas kitam prieštaraujančių lūkesčių (Nitz, 1993). Atskiri bet kurio proceso dalyviai turi skirtingus interesus, kurių tuo pat metu neįmanoma patenkinti. Vertintojams tai sukuria dilemas ar konfliktines situacijas. Bandysime išanalizuoti tokių konfliktinių situacijų struktūrą ir kai kurias joms būdingas vertybes; taip pat aptarsime kai kurias tipines konfliktines situacijas ir pateiktus pasiūlymus. Kokybės vertinimo ir vadybos kontekstas, kuriame diskutuojamos šios problemos, yra Zalcburgo universitete atliktas projektas „Studentai vertina dėstymą“ (Žr. darbo grupė „Evaluation von Lehrveranstaltungen an der Universität Salzburg“, 2002, kur pateikiama visa informacija ir sąsajos).

- *Fidelity*: here reliability; accuracy with which a characteristic is assessed with a given instrument.
- *Formative evaluation*: evaluation which is used to influence the process which has been analyzed.
- *Goals in evaluation/ education*: describes what someone aims at in a given situation.
- *Means in evaluation/ education*: describes what someone can do to achieve a goal in a given situation
- *Qualitative conflict situation*: decision situation in which the possible actions or the achievement of values are mutually exclusive (either-or)
- *Quantitative conflict situation*: decision situation in which the possible actions can be described as “more or less”, with the more one value is satisfied, the less the other can be satisfied
- *Responsibility*: (1) the force that binds you to your obligations and (2) the courses of action demanded by that force, with the force being either external (e.g., norm systems defined by other people) or internal (e.g., norm systems defended by the protagonist, conscience)
- *Responsive evaluation (model)*: important values of a responsive evaluation process are (1) transparency about all steps of the evaluation, (2) participation of all involved persons, e.g. partly in important decisions of the evaluation process, and (3) consideration of further influences, e.g. knowledge about criterion of the evaluation
- *Social reference norm*: Reference norm to evaluate students based on social comparisons, i.e., referring to the quality of the work compared to other students of a given group
- *Summative evaluation*: evaluation of the situation at a given moment in time (particularly at the end of a process)
- *Transparency*: here the protagonists have full knowledge about the goals and processes
- *Value*: something that is important for someone; the quality (positive or negative) that renders something desirable or valuable
- *Values conflict*: one person has many different but simultaneous values (intrapersonal), and between persons there may be different values (interpersonal). The different values can conflict in varying situations of action.

INTRODUCTION

People working in the intersection of two or more social networks are confronted with a multitude of expectations which often are contradictory (Nitz, 1993). The different stakeholders have different interests which cannot be fulfilled simultaneously. For evaluators or quality assessors, this creates dilemma or conflict situations. In the following we want to analyze the structure of such conflict situations and some of the typical values involved in such situations; further, we want to discuss some prototypical conflict situations and the solutions that were proposed. The quality assessment and management context used to discuss these issues is the project “Student Ratings of Teaching” at the University of Salzburg (see Task Force “Evaluation von Lehrveranstaltungen an der Universität Salzburg”, 2002, for information and links).

1 VERTYBIŲ KONFLIKTŲ SPRENDIMAS

Antagonistinės ir dileminės² situacijos – tai tokios aplinkybės, kuriomis protagonistas turi pasirinkti mažiausiai iš dviejų veiksmų, bet kad ir ką jis ar ji darytų, būtinai pažeis vieną arba kitą vertybę ar susidurs tam tikros pažeidimų normos (žr. Gastager, 1999). Terminas *antagonizmas* yra žodžių „priešingas“ ar „prieštaravimas“ sinonimas. Šis terminas atsirado dar E. Kanto laikais. Kalbant apie antagonistinius interesus ar tikslus aplinkybėmis, kuriomis protagonistas turi apsispręsti, ką daryti (žr. Thiel, 1980, p. 125ff), du interesai – I_1 ir I_2 – yra vadinami antagonistiniais, jei intereso I_1 siekimas slepia savyje intereso I_2 prevenciją, ir atvirkščiai. Tokios antagonistinės situacijos gana dažnos; žmonėms būdinga turėti keletą tuo pat metu egzistuojančių vertybių ar tikslų, kuriuos jie nori patenkinti, ir daugeliu atveju šios vertybės būna antagonistiškos.

Analizuojant tokias antagonistines ir dilemines situacijas, du dalykus reikia išskirti vertybių situacijos struktūros atžvilgiu (žr. 1 lentelę): skirtumą konflikto lygmenyje (tikslai ir priemonės) ir skirtumą konflikto rūšies atžvilgiu (kokybinis ar kiekybinis).

1 lentelė. **Konfliktų hierarchinė struktūra** (pagal Patry, 1997a)
Table 1. **A taxonomy of conflicts** (after Patry, 1997a)

		Konflikto rūšis / Type of conflict	
		Kokybinis / Qualitative	Kiekybinis / Quantitative
Konflikto vieta / Location of the conflict	Tikslai / Goals		
	Priemonės / Means		

Dviejų vertybių nesuderinamumas gali būti *tikslų* arba *priemonių* lygmenyje (Patry, 1997a). Tikslų lygmenyje konfliktą galima apibūdinti taip:

1. Asmuo siekia tikslų G_1 ir G_2 .

2. Situacija, apibūdinama G_1 – t. y. situacija, kuri susidarys, kai bus pasiektas tikslas G_1 , negali atsirasti tuo pat metu kaip ir situacija, apibūdinta G_2 .

Pateikiame įvertinimo situacijos pavyzdį. Įvertintojas respondentams, atsakinėjantiems į atviro tipo anketos klausimus, garantavo anonimiškumą, bet administracija reikalauja atskleisti kai kurių teiginių autorius. Respondentų anonimiškumo išlaikymo tikslas nesuderinamas su tapatybės atskleidimo tikslu, nes pirmasis yra logiškas pastarojo neiginys.

Antrasis konfliktas, kylantis priemonių lygmenyje, yra toks: specifinėje situacijoje žmogus turi du tikslus – G_1 ir G_2 . Tam, kad pasiektų G_1 , reikėtų atlikti veiksmą A_1 , o pasiekti G_2 reikėtų atlikti veiksmą A_2 . G_1 ir G_2 nėra vienas kitą paneigiantys, bet A_1 ir A_2 negali būti atliekami tuo pat metu. Nėra tokio veiksmo (ar veiksmų kompleks) , kuris leistų pasiekti abu tikslus G_1 ir G_2 . Pavyzdžiui, tikslai gauti patikimus duomenis (G_1) ir išleisti nedaug pinigų (G_2) logiškai nėra prieštaraujantys vienas kitam, bet aukštos

1 DEALING WITH VALUES CONFLICTS

Antagonistic and dilemma² situations are decision situations in which the protagonist has to choose between at least two actions, but whatever he or she does, he or she will necessarily break one or another value, or certain norms clash (see Gastager, 1999). The term “antagonism” is synonym of “contrary” or “contradiction”. The expression “antagonism” goes back to Kant. With respect to antagonistic interests or aims in situations in which the protagonist has to make a decision what to do (see Thiel, 1980, 125ff), two interests I_1 and I_2 are called antagonistic if the pursuit of I_1 implies the prevention of I_2 and vice versa. Such antagonistic situations are frequent; it is typical for people to have several simultaneous values or goals they want to satisfy, and in many cases these values are antagonistic.

For the analysis of such antagonistic and dilemma situations, two distinctions with respect to the structure of the situation must be made (see table 1): a distinction on the level of the conflict (goals vs. means) and one with regard to the type of conflict (qualitative vs. quantitative).

The incompatibility of two values can be either on the *goal* level or on the *means* level (Patry, 1997a). On the goal level, the conflict can be formulated as follows:

1. A person aims at the goals G_1 and G_2 .

2. The situation described by G_1 – i.e., the situation which would occur if the goal G_1 was achieved – cannot occur simultaneously with the situation described by G_2 .

An example from evaluation. An evaluator has warranted anonymity to the subjects responding to an open-ended questionnaire, but is required by the administration to reveal the identity of the author of the different statements. The goal of keeping the respondents anonymous is incompatible with the goal of revealing their identity, the former being the logical negation of the latter.

The second conflict – on the means level – is such that the person has the two goals G_1 and G_2 in a particular situation. To achieve G_1 , the action A_1 would be appropriate, while to achieve G_2 one should do A_2 . G_1 and G_2 are not mutually exclusive, but A_1 and A_2 cannot be performed simultaneously. There is no action (or set of actions) which would permit to achieve both G_1 and G_2 . For instance, the goals of valid data (G_1) and of low costs (G_2) are not logically contradictory, yet *providing* high

² Antagonistinių ir dileminų situacijų skirtumą aptarsime vėliau; žr. *kokybiniai ir kiekybiniai konfliktai*.

² For the distinction between antagonistic and dilemma situations, see below: qualitative vs. quantitative conflicts.

kokybės duomenų *pateikimas* (A_1) susijęs su daugybe darbo (A_2), kuris turi būti gerai atlyginamas.

Norint įsitikinti, ar tai tikslų konfliktas (pirmasis tipas), ar ne, reikia pažvelgti į tai, kaip, ar situacijų suderinamumas apibūdinamas pagal iškeltus tikslus; paprastai tokie aprašymai iš dalies yra tikslų apibūdinimas. Kita vertus, priemonių konfliktų metu sprendimus priimantis asmuo visų pirma galvoja apie antagonizmą; žmonės dažnai mano, kad tuo pat metu neįmanoma pasiekti visų tikslų (A_1 ir A_2 yra nesuderinami), tačiau iš tikrųjų tai visiškai įmanomas dalykas, t.y., dar yra veiksmas A_3 , apie kurį protagonistas nepagalvoja ir kuris patenkina G_1 ir G_2 (pvz., perėinant į kitą problemos lygmenį; viena galimybė būtų taikyti metakomunikaciją; plg. Watzlawick ir kt., 1969). Mėnetu atveju galbūt būtų galima rasti kitų atlygio šaltinių, ne tik pinigų: pvz., dalį darbo gali atlikti doktorantūros studentai rašydami savo daktaro disertaciją.

Antrasis skirtumas, kurį reikia aptarti, yra *kokybiniai* ir *kiekybiniai* konfliktai (žr. stulpelius 1 lentelėje). Šis skirtumas glūdi skalės, naudojamos elgsenai apibūdinti, lygmenyje. Kai kuriuos elgsenos tipus geriausia apibūdinti pasitelkiant stebėjimo sistemą, kuri suteikia duomenų nominaliajai skalei; šiuo atveju subjektas gali rinktis *arba* A_1 , ar A_2 . Mes tai vadiname *kokybiniais konfliktais*. Pavyzdžiui, įvertintojas subjektų vardus atskleidžia arba neatskleidžia. Čia kompromiso negali būti, atliekant veiksmą A_1 , mes būtinai pažeisime vertybes, susijusias su A_2 (čia nėra svarbu ar tikslo, ar priemonių lygmenyje), ir atvirkščiai. Tai vadinamoji *dilema*; tokios rūšies konfliktai paprastai sprendžiami galimybių ir vertybių teorijose (pvz., Feather, 1982; Krampen, 1986) ar moralinių sprendimų studijose (Kohlberg, 1984).

Kitokios rūšies elgseną galima apibūdinti pasitelkus rangų skalę, t.y. vartojant terminus *daugiau* ar *mažiau*. Pavyzdžiui, anketa gali būti ilgesnė ar trumpesnė. Tokiu atveju konfliktai atspindi elgsenos laipsnį intervale tarp *daugiau* ar *mažiau*, tokius konfliktus vadiname *kiekybiniais konfliktais*. Paprastai subjektai nenukrypsta į kraštutinius. Pavyzdžiui, kalbant apie anketų ilgumą kraštutinumai gali būti vienas klausimas; kita vertus, teoriškai nėra pasakyta, kokio ilgumo anketa turėtų būti. Tačiau paprastai anketos būna nuo vieno iki 20 puslapių. Pasirinkdamas tam tikrą anketos puslapių skaičių įvertintojas gali siekti kompromiso tarp respondento motyvacijos (kuo trumpesnė anketa, tuo didesnė motyvaciją ją pildyti) ir gaunamos informacijos (kuo ilgesnė anketa, tuo didesnė tikimybė, kad informacija bus visapusiškesnė). Apsisprenddamas dėl anketos ilgio įvertintojas gali apsibrėžti, kurios vertybės svarbesnės, o tai nulems anketos ilgį. Kad ir ką įvertintojas nuspręstų, jis ar ji negalės maksimaliai sutelkti dėmesio į visus rūpimus klausimus, bet turės rasti optimalų variantą ir kiek galima daugiau atsižvelgti į visas vertybes. Negalima nuspręsti apibendrintai, koks yra tas optimalus variantas (pvz., vieno ar dešimties puslapių anketa), tai priklauso nuo specifinių sąlygų ar, kitais žodžiais tariant, nuo situacijos specifiškumo (Patry, 1991); pavyzdžiui, imkite paskaitos nutraukimo metodą (Patry, 1997b), kai paskaita toliau neskaityta, o studentai klausinėjami apie tai, kas vyko per paskutiniąsias penkias minutes ar panašiai; tokiu atveju ilgesnė nei vieno puslapio anketa negali būti naudinga (bet tokį vertinimą galima pakartoti keletą kartų), o baigiamoji (*summative*) anketa gali būti ir ilgesnė. Priklausomai nuo to, kokių sąlygų reikia atsakymams (pvz., jei į anketos klausimus galima atsakinėti visą pamoką, vidutinis

quality data (A_1) requires usually much work (A_2) which must be paid for.

For the decision whether there is a goal conflict (first type) or not one can look at the compatibility of the situations described by the goals; usually, such descriptions are part of the goal descriptions. On the other hand, for a conflict of means, the antagonisms are primarily “in the head” of a decider; often people think that the simultaneous satisfaction of their aims is impossible (A_1 and A_2 are incompatible), although it would be quite possible to do so, i.e., there is an action A_3 of which the protagonist did not think and that satisfies both G_1 and G_2 (e.g., through leaving the level of the problem; one possibility would be practicing meta-communication, cf. Watzlawick et al., 1969). In the example mentioned above, maybe one can find other rewards than money – e.g., some of the work is done by graduate students as part of their doctoral dissertation.

The second distinction refers to *qualitative* vs. *quantitative* conflicts (columns in table 1). This distinction is rooted in the scale level used to describe the behavior. There are behavior types which best are described using an observation system which provides data on a nominal scale; in this case the subject can do *either* A_1 *or* A_2 . We call this “qualitative conflicts”. For instance, an evaluator reveals the names of the subjects, or he or she does not. A compromise is not possible, and doing A_1 will necessarily brake the values linked with A_2 (whether on the goal or on the means level is not relevant here), and vice versa. This is called “dilemma”, and this kind of conflict is typically dealt with in expectancy-values theories (e.g., Feather, 1982; Krampen, 1986) or in moral judgment studies (Kohlberg, 1984).

Other types of behavior can be described on an ordinal level, i.e., with terms like “more” and “less”. For instance, a questionnaire can be longer or shorter. The conflicts then are conflicts with respect to the degree of a behavior, between “more” and “less”; we call them “quantitative conflicts”. Usually the subjects do not perform the extremes. For instance, with respect to the length of the questionnaires, at the lower end, one could use only one item; on the other end, there is no theoretical upper limit of the length of a questionnaire. However, in practice one will use questionnaires of between one page and 20 pages. By choosing a certain number of pages the evaluator can make a compromise between keeping the responders’ motivation (the shorter the questionnaire, the better the motivation) and getting information (the longer, the more information). By deciding on the length of the questionnaire the evaluator can decide to what degree to take into account for the different values associated with questionnaire length. Whatever the evaluator decides he or she will not maximally respect the values at stake, but an optimum is possible in which the values are achieved as much as possible. Where this optimum lies (e.g., one page or ten pages for a questionnaire) cannot be decided in general, but depends on the particular conditions or, in other term, is situation specific (Patry, 1991); for instance, with the lesson interruption method (Patry, 1997b) in which the lesson in the classroom is interrupted at a given time and the students are questioned about what happened in the last five minutes or so, a questionnaire of more than one page will not work (but one can repeat the assessment several times), whereas a summative questionnaire can be much longer, depending

studentas ją turėtų užpildyti maždaug per 40 minučių). Šiuo atveju negalime kalbėti apie dilemą, nes negalima išskirti tik dviejų galimybių, labiau tiktų kalbėti apie antagonistines situacijas.

Tiek bendro pobūdžio įvertinimui, tiek specifiniam įvertinimui labai svarbu, ar konfliktas kokybinis (dilema), ar kiekybinis (antagonistinis), nes galimi konflikto sprendimo būdai yra visiškai skirtingi. Tikroje dileminėje situacijoje sprendimo priimti neįmanoma atsižvelgiant į visas su konfliktu susijusias vertybes ir interesus, o kiekybinio konflikto atvejais kompromisas gali būti pasiektas, kai tam tikru laipsniu į tas skirtingas vertybes atsižvelgiama. Smulčiau apie šį skirtumą ir jo implikacijas apskritai skaitykite Patry (1997a); šiame straipsnyje dėmesį sutelksime į šio skirtumo taikymą įvertinimams ir kokybės vertinimui.

2 KONFLIKTUOJANČIOS VERTYBĖS ĮVERTINIME

Kaip minėta ankstesniame skyriuje, konfliktinėmis aplinkybėmis susiduria keletas vertybių, interesų ir tikslų. Visame švietimo įvertinimo procese atsakomybė laikoma pagrindine vertybe. Smith ir Glass (1987, p.47) siūlo atliepiamąjį (*responsive*) įvertinimo modelį, kuriame akcentuojama atsakomybė, skaidrumas ir teisingumas: „Tam, kad atliktų atliepiamąjį įvertinimą, įvertintojas sukuria stebėjimo ir derybų planą. Jis planuoja, kuriuo metu įvairūs asmenys stebės programos įgyvendinimą“ (Smith ir Glass, 1987, p. 47). Šiuo metu tokie reikalavimai dažniausiai taikomi kiekvienos švietimo situacijos įvertinimui, pvz., „atsakomybė už dėstymo sritį“ yra pagrindinis reikalavimas „Etiniuose Amerikos švietimo tyrimų asociacijos standartuose“ (AERA, 1992). Bet įvairių asmenų, suinteresuotų institucijos veikla, atsakomybė gali reikalauti netapusių veiksmų; tad kitos vertybės, o ne vien atsakomybė, tampa svarbios, vadinasi įvertinimas visada sukuria dilemą ar antagonistinę situaciją (žr. Cronbach, 1985).

Kokios vertybės yra aktualiausios? Metaįvertinimo tyrimas apimantis įvertintojo dilemas ir antagonizmus remiantis patyrusių įvertintojų apklausa (Gastager ir Patry, 2001), išskėlė šešias vertybių sritis: etines, metodines, socialines ir interaktyviasias, įstatymines, ekonomines ir asmenines; visas jas galite rasti literatūros apie įvertinimą šaltiniuose, pvz., Cronbach (1985, 1987), įvertinimo etinėse rekomendacijose (pvz., AERA 1992) ir įvertinimo standartuose (pvz., Deutsche Gesellschaft für Evaluation, 2002; Paschon ir Thonhauser, 2002). Įvertinimui reikalingos vertybių sritys ir vertybių, kurios išryškėja įvertintojų konfliktinėse situacijose, pavyzdžiai pateikti lentelėje 2.

1. Pirmoji sritis – *etinės vertybės* grindžiamos etiniais pasiteisinimais, apibūdintos „Amerikos švietimo tyrimų asociacijos etiniuose standartuose“ (AERA, 1992). Tokių vertybių pavyzdžiai – sąžiningumas ir skaidrumas (pvz., įvertinimo priemonių ir svarbiausių problemų aiškinimas tiriamiesiems). Kita standartinė vertybė, minima daugelyje įvertinimo tyrimų – anonimiškumas; tačiau kartais žmonės pageidauja atverti savo asmeninius duomenis viešai (pvz., Shulman, 1990).

2. Antroji vertybių sritis susijusi su *metodais*: bet kuris vertinimas turi būti patikimas, vertinimo projektas –

on the conditions of responding (e.g., if a full lesson is reserved for answering the questionnaire, the average subject should be able to fill it in within about 40 minutes). One cannot speak of a dilemma here because there are not only two possibilities; it is more appropriate to speak of antagonistic situations.

Whether a conflict is qualitative (dilemma) or quantitative (antagonistic) is very important in general and for evaluation in particular because the possible solutions are completely different. In a true dilemma situation a solution taking into account all involved values and interests is impossible, while in situations with a quantitative conflict a compromise can be conceived in which the different values can be taken into account to a certain degree. For more details on this distinction and its implications in general see Patry (1997a); here we will concentrate on its application to evaluations and quality assessment.

2 CONFLICTING VALUES IN EVALUATION

As presented in the previous chapter in conflict situations several values, interests or goals clash. In the complete process of an educational evaluation a central value is responsibility. With Smith and Glass (see 1987, 47) one should consider a responsive evaluation model which includes responsibility, transparency, and fairness: “To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges various persons to observe the program” (Smith & Glass, 1987, p. 47). Nowadays in every educational evaluation setting these demands are put into practice; e.g., “responsibility to the field” is a Guiding Standard of the Ethical Standards of the American Educational Research Association (AERA, 1992). But responsibility with respect to different stakeholders may mean different actions, and other values than responsibility may be important, hence evaluation always has a dilemma or antagonistic structure (see Cronbach, 1985).

What values are at stake? In a meta-evaluation study of evaluators' dilemmas and antagonisms based on interviews with experienced evaluators (Gastager & Patry, 2001), six areas of values were identified: ethical, methodical, social and interactive, legal, economic, and personal values; all of them can be found at some place in the literature on evaluation, e.g., in Cronbach (1985, 1987), in the ethical guidelines (e.g., AERA 1992) and standards for evaluation (e.g., Deutsche Gesellschaft für Evaluation, 2002; Paschon & Thonhauser, 2002). In table 2, these areas and examples for the values involved in evaluators' conflict situations and which are quite typical for the field are listed.

1. The first area, *ethical values*, includes values based on ethical justification and refers mostly to the Ethical Standards of the American Educational Research Association (AERA, 1992). Examples for values are honesty and transparency (e.g., explaining the instruments and the central issues of the evaluation to the investigated persons). A further value is anonymity, which is a standard in most evaluation studies; however, sometimes the persons prefer to open their personal data to the public (e.g. Shulman, 1990).

2. The second values area concerns the *methods*: any assessment needs to be valid, the design must be such

2 lentelė. **Vertybių sritys ir jų pavyzdžiai** (adaptuota iš Gastager ir Patry, 2001)
Table 2. **Values areas with examples** (adapted from Gastager & Patry, 2001)

VERTYBIŲ SRITIS / VALUES AREA	PAVYZDŽIAI / EXAMPLES
1 Etinės vertybės	a. Sąžiningumas ir skaidrumas; b. anonimiškumo apsauga; c. teisingumo būtinybė; d. įvertinamos institucijos noras gauti informaciją; e. įvertinamų asmenų apsauga.
1 Ethical values	a. honesty/ transparency; b. protection of the anonymity; c. obligation for truthfulness; d. wish for information of the evaluated institution; e. protection of the evaluated persons.
2 Metodinės vertybės	a. objektyvumas, patikimumas, atitinkami metodai; b. aprėptis; c. priemonių tinkamumas; d. matavimo objektyvumo siekis; e. dalyvavimo laipsnis; f. hipotezės patvirtinimas; g. įvertintojo profesionalumas.
2 Methodical values	a. objectivity/ reliability/ validity, appropriate methods; b. bandwidth; c. practicability of the instruments; d. ideal of an uninfluenced measurement; e. degree of participation; f. confirm the hypothesis; g. professionalism of the evaluator.
3 Socialinės ir interaktyvios vertybės	a. noras gauti informaciją apie rezultatus; b. viešumas; c. atskirų įvertinamų grupių tikslų skirtingumas; d. intervencijos efektyvumo lūkesčiai; e. įvertintojo ir įvertinamos institucijos reikalavimas dalyvauti ir būti lygiateisiu vertinimo proceso nariu; f. įvertinamų asmenų (pvz., lektorių, dėstytojų) profesionalumas.
3 Social and interactive values	a. wish for information about results; b. publicity; c. different aims of the different evaluated groups (plurality of aims); d. expectations concerning the effectiveness of the intervention; e. claim for participation and emancipation between evaluator and evaluated institution; f. professionalism of the evaluated persons (e.g. lecturer, teacher).
4 Įstatyminės vertybės	a. įstatymų leidėjo profesionalumas; b. noras adaptuoti įvertinamąjį objektą; c. noras gauti informaciją apie įvertinamą instituciją; d. grįžtamojo ryšio apie rezultatus reikalavimas; e. įvertinimu paremtas sprendimo priėmimas.
4 Legal values	a. professionalism of the legislator; b. wish for adaptation of the evaluation object; c. wish for information of the evaluated institution; d. claim for feedback concerning the results; e. decisions based on evaluation.
5 Ekonominės vertybės	a. kaštai; b. įvertintojų pastangų išvengimas; c. įvertinimo išlaidos; d. įvertinimo apimtis (pvz., trukmė ir t.t.); e. rezultatų naudingumas.
5 Economic values	a. costs; b. avoiding efforts of the evaluators; c. expense for the evaluation; d. extent of the evaluation (e.g. concerning time etc.); e. (practical) usefulness of the results.
6 Asmeninės vertybės	a. įvertinamų asmenų privataus gyvenimo apsauga; b. „tiesioginė pagalba“ tiriamiems asmenims; c. domėjimasis tiriamuoju objektu; d. nerimas dėl rezultatų išspausdinimo; e. specifinis grįžtamojo ryšio poveikis individui.
6 Personal values	a. protection of the private sphere of the evaluated persons; b. “direct help” for investigated persons; c. interest in the investigation object; d. anxiety concerning the publication of the results; e. aimed/concrete feedback for the individual.

leidžiantis atmesti visas susikertančias hipotezes (Campbell, 1969), duomenų analizėje turi nebūti klaidų ir t. t. Ypač svarbus klausimas šioje srityje yra įvertintojų ir švietimo institucijų profesionalumas.

3. Trečioji sritis – *socialinės ir interaktyviosios* vertybės, apimančios įvairių suinteresuotų grupių pageidavimus ir lūkesčius. Šių grupių interesai gali būti visiškai skirtingi, antagonistiniai ar net nesiderinantys. Kai kurios suinteresuotos grupės (pvz., administracija) gali norėti gauti specifinę informaciją; kiti (pvz., personalas) gali nenorėti šios informacijos atskleisti. Tam tikrų suinteresuotų grupių informavimas artimai susijęs su bendru švietimo institucijos reikalavimu domėtis grįžtamoju ryšiu, nes iš grįžtamojo ryšio gauname informaciją apie tai, kokių rezultatų pasiekta įdėjus tam tikrų pastangų (žr. Slavin, 1986, p. 374). Tačiau tai nereiškia, kad kiekviena suinteresuota grupė turėtų gauti bet kokią juos dominančią informaciją. Kitas pavyzdys, atskleidžiantis prieštaringas suinteresuotų grupių vertybes – dažnai pasitaikantis administracijos keitinimas ką nors keisti, kai dirbantieji nieko keisti nenori.

4. Ketvirtoji sritis – *teisinės vertybės*; tai visos tos vertybės, kurios pateikiamos įstatymuose ir tie, kurių pareiga priversti tuos įstatymus vykdyti. Šiuos paprastai gana apibendrintus reikalavimus galima įvairiai interpretuoti ir taikyti specifinei situacijai; įvertintojo interpretacijos kartais skiriasi nuo įstatymo administratorių interpretacijų.

5. Penktoji sritis – *ekonominės vertybės*; tai vertybės, susijusios su įvertinimo finansiniais ir personalo aspektais ar laiko resursais, skirtingų suinteresuotų grupių pastangomis ir išlaidomis, išoriniais šaltiniais, medžiagomis, technine ir programine įranga ir t. t.

6. Paskutinioji – sritis – „*asmeninių vertybių*“ sritis; čia įeina visos suinteresuotų asmenų, įvertintojų ir t. t. asmeninės vertybės. Vertybių pavyzdžiai: privataus įvertintojų gyvenimo apsauga ar asmeniniai įvertintojo interesai ir t. t.

Ši vertybių struktūra yra sugalvota analitiniam tikslams. Sričių eiliškumas, kaip parodyta 2 lentelėje, nereiškia jokio prioriteto ranguojant; kiekvienas atskiras įvertintojas pats nusprendžia, kurioms vertybėms teikti prioritetą duotoje konfliktinėje situacijoje. Konfliktai gali iškilti vienoje vertybių srityje (pvz., tarp metodinių vertybių: aprėpties ir patikimumo dilema; Cronbach, 1970, p. 179 ff., dar žr. 4.1 poskyris; konfliktai tarp įvairių suinteresuotų grupių, žr. trečiąją vertybių sritį) arba tarp skirtingų vertybių sričių (pvz., metodinės ir ekonominės vertybės).

Visai įmanomas dalykas, kad kai kurios vertybės priklauso dviem ar kelioms vertybių sritims. Visos išvardytos vertybės svarbios, tačiau kuriai sričiai teikti pirmenybę specifinėje situacijoje, turi nuspręsti įvertintojas ar kiti suinteresuoti asmenys, atsižvelgdami į konkrečias sąlygas; pasak Cronbach (1987, p. 321), „įvertinimas yra menas“, todėl „nėra vienintelio geriausio įvertinimo plano“.

3 STUDENTŲ DĖSTYMO REITINGAVIMAS ZALCBURGO UNIVERSITETE

Prieš pradėdami taikyti struktūrinę hierarchijos principą (taksonomiją) tyrimui, kurio metu Zalcburgo universiteto studentai reitingavo dėstymą, turėtume pirmiausiai apibūdinti vertinimo projektą ir kontekstą. 1993 m. Austrijo-

that plausible rival hypotheses are ruled out (Campbell, 1969), the analysis of the data has to be correct, and so on. A particularly important issue is professionalism especially of the evaluators and the educational institutions.

3. The third area is on *social and interactive* values and concerns requests and expectancies from the different stakeholder groups. These groups may have different and antagonistic or mutually exclusive interests. For instance, some stakeholders (e.g., authorities) may wish specific information; others (e.g., personnel) may not want to reveal it to them. Informing certain stakeholders is closely related to the general demand in educational settings for feedback, since “feedback” means information on the results of one’s efforts (see Slavin, 1986, p. 374). However, this does not mean that every stakeholder should get any information he or she might be interested in. Another example of opposite stakeholder values which can often be found is the intention of the administration to change something whereas the employees do not want to change.

4. Fourth, there is the area of *legal values*, which includes all values provided by the law and by those who are supposed to enforce it. These usually rather general requirements are subject to interpretation and application to the specific situation; the evaluator’s interpretation sometimes differs from the interpretation by the administrators of the law.

5. The fifth area, *economic values*, concerns all the values connected with the financial and personnel background of the evaluation or the time resources, including the costs, the efforts to be spent by the different stakeholders, external resources, material, hard- and software, etc.

6. Finally, the “*personal values area*” deals with all the intrapersonal values of the stakeholders, the evaluators, etc. Examples for values are protection of the private sphere of the evaluated persons or the personal interests of the evaluator in the evaluation, etc.

This taxonomy of values is conceived for analytical purposes. The order of the areas in table 2 is not meant to present some priority ranking; rather, it is the responsibility of the individual evaluator to decide which values to prefer in a given conflict situation. Conflicts can be within one single values area (e.g., within methodical values: the bandwidth-fidelity dilemma, Cronbach, 1970, pp. 179ff., see below, section 4.1; conflicts between different stakeholder groups, see the third area) or between values areas (e.g., methodical vs. economic values).

It is quite possible that a certain value fits into two or more of the areas. All values mentioned seem quite important, yet which one is to prefer in the particular situation is a decision which the evaluator or other stakeholders have to take considering the actual conditions; as Cronbach (1987, p. 321) says as general advice: “Evaluation is an art” and “(t)here is no single best plan for an evaluation.”

3 STUDENT RATINGS OF TEACHING AT THE UNIVERSITY OF SALZBURG

Before applying the taxonomy to the student ratings of teaching at the University of Salzburg, this evaluation project and its context need to be described. In Austria, a new law about the organization of universities

je buvo išleistas naujas įstatymas apie universitetų reorganizavimą (UOG 1993), kuris Zalcburgo universitete buvo įgyvendintas 1997 m. (šiuo metu minėtas įstatymas jau pakeistas). 18-ame šio įstatymo paragrafe numatyta, kaip organizuoti tyrimų ir dėstymo įvertinimą universitete. Be kitų nurodymų buvo pateiktas ir detalus reglamentas, kaip organizuoti įvertinimo procesą, kai studentai sudarinėja dėstymo reitingus; šie reglamento nurodymai apibūdinti daug išsamiau nei kuri nors kita įvertinimo sritis (Patry, 2002). Vykdydamas šį įsakymą, Zalcburgo universiteto studijų prorektorius inicijavo darbo grupės suformavimą, kurios tikslas buvo tokį įvertinimą atlikti. Įvertinimo koncepcija (Bauman ir kt., 1999) buvo pateikta universiteto bendruomenei viešai aptarti, 2000 vasarą atliktas žvalgomasis tyrimas, o du visapusio pasiekimų įvertinimo ciklai vyko 2001 vasarą ir 2001–2002 žiemą. Buvo planuoti ir vėlesni įvertinimai, tačiau šiuo metu įgyvendinamas naujas įstatymas, kur įvertinimas ne toks griežtas; pastaruoju metu jau akcentuojamos kitos, ne įvertinimo, sritys, o įvertinimo srityje daroma ne daug kas.

3.1. ĮVERTINIMO SĄLYGOS

Pirmoji darbo grupės užduotis buvo sukurti koncepciją (Baumann ir kiti, 1999) ypatingą dėmesį kreipiant į šiuos klausimus:

- Įvertinimas, kai studentai sudarinėja dėstymo reitingus, nėra vienintelis dėstymo įvertinimo elementas. Reikia įvertinti pažymių studentams rašymą, testavimo sistemą (ypač atsižvelgiant į sąžiningumą ir patikimumą), tai, ar studentai sugebėjo išmokti tuos dalykus, kuriuos turėjo išmokti ir t.t.
- Įvertinimo rezultatai yra prasmingi, kai įvertinimas naudojamas kaip kokybės vadybos priemonė.
- Nepaisant to, kad įstatyme prioritetas teikiamas suminiam įvertinimui, darbo grupė teikė pirmenybę formuojamajam įvertinimui, t.y. kokybės vertinimui integruoti į dėstymo proceso tobulinimą.
- Darbo grupė ypač akcentuoja konfidencialumą ir duomenų saugumą, nes duomenys – kaip ir daugumoje įvertinimo kontekstų atvejų – yra labai subtilus veiksnys; viešai neturi būti skelbiama jokia informacija apie atskirų kursų ar dėstytojų įvertinimą, pateikiama gali būti tik apibendrinta informacija. Tai reiškia, kad neketinama sudarinėti nei dėstytojų, nei kursų reitingų, nei atlikti kokio kito panašaus vertinimo – tai būtų nesuderinama su formuojamu šio metodo pobūdžiu.
- Praktinio pobūdžio priežastys lėmė tai, kad atsakymai į anketos klausimus buvo pildomi raštu ir naudojamas popierinis anketos variantas; kompiuterio anketos varianto atsisakyta darant prielaidą, kad tai gali sąlygoti žemą respondentų aktyvumo laipsnį ir sukelti kitų nepatogumų.

Šie principai buvo pristatyti universiteto bendruomenei, bet prieš pradėdant įvertinimą nebuvo sulaukta didelės dėmesio dėl būsimos įvertinimo pobūdžio, reakcijos sulaukta tik pasibaigus įvertinimo procesui.

Darbo grupė parengė specialaus pobūdžio anketą, kurios čia neaptarsime (žr. Lengenfelder ir kt., 2001; Lengenfelder, 2003) ir atliko žvalgomąjį tyrimą ($n=4887$, iš viso 296 kursai); patikimumui patikrinti buvo atlikta daugybinių kintamųjų ir daugybinių metodų analizė (Multitrait-Multimethod-Analysis) (Campbell ir Fiske, 1959) (ci-

UOG 1993) was decided in 1993 and implemented in the University of Salzburg in 1997 (in the meantime this law has been replaced by another one). In paragraph 18, the law stipulated how evaluations in research and teaching must be done in the universities. Among others, detailed regulations about student ratings of teaching were given, being much more specific than for any other domain of evaluation (Patry, 2002). Consequently, the vice-rector for teaching at the University of Salzburg initiated a Task Force with the aim to implement such an evaluation. A concept for the evaluation (Baumann et al., 1999) was submitted to the University public for discussion, a pilot study was performed in summer 2000 and two series of full assessments were done in summer 2001 and in winter 2001-2002. Further assessments were planned, however in the meantime a new law has been implemented which is not so strict with respect to evaluations; currently, the focus is on other issues than evaluation, and little is being done in this regard.

3.1 THE CONDITIONS OF THE EVALUATION

The first activity of the Task Force was to write a concept (Baumann et al., 1999) with, in particular, the following issues:

- The student ratings of teaching are not the sole element of evaluation of teaching. One should also evaluate the grading and testing system (particularly with respect to fairness and validity), whether the students learned what they were supposed to learn, etc.
- An evaluation makes sense only if it leads to consequences, i.e., when it is used as means for quality management.
- Despite the fact that the law seems to prefer summative evaluations, the Task Force preferred formative evaluations, i.e. the integration of the assessments within a process of improvement.
- The Task Force put particular emphasis on confidentiality and data protection, because the data – as in most evaluation contexts – are very sensitive; especially no information about individual courses or teachers, but only aggregated data are to be made accessible to the public. This means also that no ranking of teachers or courses or of any other kind is intended – such rankings would also be incompatible with the formative aim of the approach.
- For practical reasons, a paper-pencil response mode for the questionnaire was used; an online questionnaire was rejected because of the assumed low response rate and further disadvantages.

These principles were submitted to the University community, but ahead of the evaluation there was little interest for a discussion of them – only after the evaluation was implemented there were reactions.

The Task Force then developed a questionnaire with special features which cannot be discussed here (see Lengenfelder et al., 2001; Lengenfelder, 2003, for details) and performed a pilot study ($n=4887$ for 296 courses) with a Multitrait-Multimethod-Analysis (Campbell & Fiske, 1959) for validation (reported in Lengenfelder et al., in press). After some necessary adaptations the questionnaire was used in the summer of 2001 ($n=20228$ for 1176 courses)

tuojama iš Lengenfelder ir kt. spaudos). Atlikus reikiamus patobulinimus, anketa buvo platinama 2001 m. vasarą (n=20228; iš viso 1176 kursai), o po to 2001–2002 m. žiemą (n=20062; iš viso 1902 kursai) visame universitete. Analizės rezultatų čia nepateiksim (juos galima rasti Lengenfelder, 2003), bet atlikime kai kurių prieštaravimų analizę siekdami iliustruoti skirtumus, pateiktus 1 ir 2 poskyriuose.

3.2. SUINTERESUOTIEJI ASMENYS IR JŲ VERTYBĖS

Į bet kurį įvertinimo kontekstą įtraukiamos įvairios visuomenės grupės, turinčios skirtingas vertybes; šios diskusijos tikslu terminą *suinteresuotieji asmenys* vartosime laisvai, įtraukdami ir tas grupes, kurios paprastai su minėtu terminu nesiejamos. Šiuo atveju galime išskirti kelias suinteresuotų asmenų grupes pagal jų turimas vertybes (žr. 3 lentelę), tos vertybės smarkiai susipynusios su vertybių sritimis, pateiktomis 2 lentelėje:

1. *Vyriausybė* (federalinė administracija), būtent Austrijos federalinė švietimo, mokslo ir kultūros ministerija (atsakinga už universitetus; <http://www.bmbwk.gv.at/start.asp>) ir jos atstovai – įstatymai, jų vykdymo taisyklės ir federaliniai administratoriai. Pagal veikiančius įstatymus, galima išskirti šias vertybes (taip pat žr. 2 lentelės vertybių skiltį – 4):

a) federaliniai administratoriai siekia patikrinti dėstytojų universitete efektyvumą ir veiksmingumą ir suteikti pagrindą kokybės vadybos sprendimams; be visa kita, sprendimai dėl dėstytojų (įdarbinimas, kėlimas pareigoje, kiti su karjera susiję sprendimai, asmeninio tobulėjimo priemonės ir t. t.) privalo remtis konkretais asmens dėstyto reitingavimu, kurį atlieka studentai. Tad jei dviejų sistemingų vertinimų metu studentai dėstytoją įvertina neigiamai, turi būti imamasi priemonių dėstyto kokybei pagerinti.

b) jie siekia gauti informacijos apie dėstytojų universitete *status quo*.

c) jie siekia priversti laikytis įstatymo, pagal kurį studentai turi ranguoti dėstytoją.

2. *Universiteto administracija*, įskaitant atsakingas pareigas universitete užimančius pareigūnus – rektorius ir dekanus, prorektorius ir prodekanus, kurie atsakingi už dėstytojų kokybę, katedrų vedėjus, programų komitetų vadovus ir t. t. Visi jie kelia savo reikalavimus ir užduotis, bet jų išteklių gana riboti (žr. 5 skilties vertybes 2 lentelėje). Jų bendrosios vertybės yra šios:

a) jie nori, kad įvertinimo procesas būtų kokybiškas, nors dauguma jų nėra nei įvertinimo nei kokybės vadybos specialistai;

b) jie atlieka daugybę kitų papildomų pareigų (6 skiltis 2 lentelėje: papildomosios asmeninės vertybės).

Įvertinimo ir kokybės vadybos atžvilgiu jie paprastai laikosi šių nuostatų:

c) nori pagerinti institucijos, už kurią jie atsakingi, veiklos kokybę;

d) nori, kad būtų laikomasi įstatymų (4 skiltis 2 lentelėje ir anksčiau minėtas 1a).

3. *Dėstytojai, kurių kursai yra įvertinami*, yra asmenys, labiausiai suinteresuoti įvertinimu, kadangi jų

and in the following winter of 2001-2002 (n=20062 for 1902 courses) for the full University. We are here not reporting the analysis of the questionnaire results (which can be seen in Lengenfelder, 2003); rather we will perform an analysis of some of the antagonisms to illustrate the distinctions made above in sections 1 and 2.

3.2. THE STAKEHOLDERS AND THEIR VALUES

In any evaluation endeavor, many stakeholders with different values are involved; for the purpose of this discussion, we take the term “stakeholder” loosely and include also some groups which commonly are not listed among the stakeholders. In the present case, we can distinguish the following stakeholder groups with according values (see table 3), the latter having a large overlap with the values areas of table 2:

1) *The government* (the federal administration) is, in particular, the Austrian Federal Ministry for Education, Science and Culture (which is responsible for the universities; <http://www.bmbwk.gv.at/start.asp>) and its representatives, and it includes the laws, the rules for its execution, and the federal administrators. According to the laws in action, the following values can be identified (see also values area 4, table 2):

a) The federal administrators want to check effectiveness and efficiency of University teaching with the aim of providing a foundation for quality management decisions; among others, decisions on teaching personnel (hiring, promotion, other career decisions, personnel development measures, etc.) must be based on the student ratings of teaching of the respective person. Further, if a teacher has been rated negatively by the students for two consecutive assessments, measures to increase the teaching quality have to be taken.

b) They want information about the status quo in teaching at the universities.

c) They want to enforce the law with regard to student ratings on teaching.

2) *The University administration* includes the officials who have a responsible position within the University, beginning with the rector and the deans, but also the vice-rector and vice-deans responsible for teaching, the heads of departments, the heads of program commissions, etc. All these people have their requirements and tasks, but also only a limited volume of resources (values area 5, table 2). They have the following general values:

a) They want to provide a qualitatively good evaluation process although most of them are not specialists in evaluation or quality management, and

b) they have many additional duties (area 6 in table 2: additional intrapersonal values).

With respect to evaluation and quality management, they have typically the following values:

c) They want to increase the quality of the institution for which they are responsible.

d) They want to enforce the law (area 4 in table 2 as well as 1a above).

3) *The teachers whose courses were evaluated* are the people most directly concerned by the evaluation since the ratings are on their teaching and the decisions of the administration have a direct influence on their work

3 lentelė. **Suinteresuoti asmenys bei jų (numanomos) vertybės studentams reitinguojant dėstyją**
Table 3. **Stakeholders and their (presumed) values in the student rating of teaching**

SUINTERESUOTIEJI ASMENYS / STAKEHOLDERS	NUMANOMOS VERTYBĖS / PRESUMED VALUES
1 Vyriausybė	a) patikrinti dėstyimo universitete efektyvumą ir veiksmingumą siekiant priimti kokybės vadybos sprendimus; b) gauti informaciją apie dėstyimo universitete <i>status quo</i> ; c) priversti laikytis įstatymo dėl studentų atliekamų dėstyimo reitingų.
1 Government	a) check effectiveness and efficiency of University teaching to make quality management decisions; b) be informed about the status quo in teaching at the universities; c) enforce the law with regard to student ratings on teaching.
2 Universiteto administracija	a) įvertinimą atlikti gerai nepaisant žinių apie įvertinimo principus trūkumo; b) atlikti kitas pareigas; c) pagerinti darbo institucijoje kokybę; d) priversti laikytis įstatymų.
2 University administration	a) provide good evaluation despite lack of knowledge in evaluation principles; b) perform other duties; c) increase the quality of the institution; d) enforce the law.
3 Dėstytojai	a) priimti ir palaikyti įvertinimus; b) atitinkamai sąžiningi ir, jei galima, teigiami studentų reitingai apie dėstyją; c) jei reitingai neigiami – jokio viešo duomenų naudojimo ir jokių bendrų diskusijų; d) aiškus grįžtamasis ryšys; e) sumažinti iki minimumo pastangas ir laiką; f) pagerinti dėstyją neįdedant per daug pastangų.
3 Teachers	a) accept and support evaluations; b) appropriate, fair, and if possible positive student ratings; c) if the ratings are negative: no general access to the data and no general discussion; d) understandable feedback; e) minimize effort and time; f) improve teaching without putting much effort into it.
4 Studentai	a) jei reikia, pagerinti dėstyimo kokybę; b) sąžiningumas vertinant dėstyją; c) nepadaryti žalos dėstytojui; d) sumažinti iki minimumo pastangas ir laiką; e) pakartotinai neatsakinėti į tą pačią anketą.
4 Students	a) increase teaching quality if necessary; b) honesty in rating teaching; c) no harm to the teacher; d) minimize effort and time; e) not responding to the same questionnaire repeatedly.
5 Mokslinė bendruomenė	žr. 2 vertybių skiltį, 2 lentelė
5 Scientific community	See values area 2, table 2
6 Įvertintojai	a) pagerinti dėstyimo universitete kokybę; b) įvertinimas kaip tyrėjo pastanga; c) įstatymus taikyti remiantis savomis interpretacijomis; d) sumažinti iki minimumo pastangas ir kaštus.
6 Evaluators	a) improve the teaching quality at the University; b) evaluation as research endeavor; c) apply the law according to their own interpretation; d) minimize their efforts and costs;

darbo reitingai ir administracijos sprendimai daro tiesioginę įtaką jų veiklai ir karjerai. Tai labai heterogeniška grupė; galima išskirti šias jų vertybes:

- a) jie palaiko įvertinimą, ir sutinka, kad jų darbas būtų įvertinamas.
- b) jie atitinkamai tikisi sąžiningo ir, jei įmanoma, teigiamo studentų vertinimo.
- c) jei vertinimas neigiamas, jie nenori, kad apie tai būtų viešai skelbiama ir diskutuojama.
- d) jie nori sulaukti tokio grįžtamojo ryšio, kurį jie supranta.
- e) jie nenori skirti per daug laiko ir dėti per daug pastangų įvertinimo procesui.
- f) jie nori tobulinti savo dėstymą neįdedant pernelyg daug pastangų.

Galbūt šios vertybės nėra būdingos visiems dėstytojams, tačiau mūsų patirtis rodo, kad jos būna gana tipiškios.

4. *Studentai* taip pat tiesiogiai susiję asmenys, tačiau kitokiu būdu. Jie taip pat sudaro heterogenišką grupę, kurios pagrindinės vertybės, manytume, yra šios:

- a) jie siekia, kad dėstyimo kokybė gerėtų;
- b) jie nori įvertinti dėstyimo kokybę kaip galima sąžiningiau;
- c) jie nenori padaryti žalos dėstytojui;
- d) jie nenori sugaišti daug laiko anketų pildymui;
- e) jie nenori iš naujo pildyti tas pačias anketas, vertindami kelis kursus.

5. *Mokslinė bendruomenė* – tai įvertinimo ekspertai, kurie dažniausiai yra įvertintojų kolegos. Jų vertybės atsispindi atlikto įvertinimo kritikoje. Šios grupės tikslas – apginti įvertinimo teorinius ir praktinius aspektus, pasiekti tokį įvertinimą, kuris atitiktų esamą padėtį. Jų vertybės pateiktos metodinėje (2 skiltis, 2 lentelė) ir etinėse skiltyse (1 skiltis).

6. *Įvertintojai* yra darbo grupės nariai: keturi universiteto dėstytojai, savanoriai, dalyvaujantys šiame darbe, vienas universiteto darbuotojas (turintis psichologijos daktaro laipsnį) ir vienas studentas, pasamdyti šiam tikslui. Įvertinimas buvo tik dviejų pastarųjų asmenų pareiga, o du universiteto dėstytojai taip pat buvo ir universiteto administracijos nariai (3 grupė). Darbo grupės vertybės aprašytos koncepcijoje (žr. aukščiau, poskyris 3.1). Be to, svarbios ir šios vertybės:

- a) darbo grupė siekia pagerinti dėstyimo universitete kokybę;
- b) įvertintojams šis projektas atveria *tyrimo* galimybes su visomis iš to išplaukiančiomis pasekmėmis: patikrinti hipotezes ir teorijas, gauti apibendrinančių žinių ir paskleisti jas mokslinei bendruomenei, pvz., publikacijų pavidalu;
- c) įstatymai taikomi atsižvelgiant į įvertintojų interpretacijas, kurios gali skirtis nuo standartinių interpretacijų, bet išlikti tarp galimų interpretacijų;
- d) visi darbo grupės nariai labai užsiėmę, todėl stengiasi sumažinti savo pastangas iki minimumo.

Šios vertybės tėra tik spėjamos, nes nėra atlikta kokių nors tyrimų joms įvertinti. Nepaisant to, sąrašas atrodo gana pagrįstas, tačiau turime žinoti, kad šis sąrašas nėra galutinis ar išsamus, kitos papildomos vertybės taip pat gali būti svarbios.

and career. This is a very heterogeneous group; nevertheless some values can be assumed:

- a) They accept and support evaluations, including being evaluated.
- b) They want appropriate, fair, and if possible positive student ratings.
- c) If the ratings are negative, they do not want them to be generally known and discussed.
- d) They want a feedback that they understand.
- e) They do not want to spend much time and effort for the evaluation.
- f) They want to improve their teaching without putting much more effort into it.

These values may not be applicable to all teachers, but according to our experience they are fairly representative.

4) *The students* are also directly concerned, though in a very different way. They are also a heterogeneous group, whose most important values are, as we assume:

- a) They hope for an increase in the teaching quality if necessary.
- b) They want to rate the teaching quality as honestly as possible.
- c) They do not want to harm the teacher.
- d) They do not want to spend much time in responding to questionnaires.

e) They do not want to respond to the same questionnaire again and again, but with respect to different courses.

5) *The scientific community* consists in experts in evaluation, particularly colleagues of the evaluators. Their values will become manifest in critiques of the evaluation as performed. This group is introduced here to defend the values of evaluation research and practice, with the focus on achieving an evaluation according to the current state of the art. The values are those presented above in the methodical (area 2, table 2) and ethical domains (area 1).

6) *The evaluators* are the members of the Task Force: four University teachers who volunteered for this work, one University employee (with a PhD in Psychology), and a student engaged for this purpose. Only the latter two had the evaluation among their duties, and two of the University teachers were also members of the University administration (group 3). The values of the Task Force are described in the concept (see above, section 3.1). In addition the following values are important:

- a) The Task Force wants to improve the teaching quality at the University.
- b) It considers the project as *research* endeavor with all its implication: among others, to test hypotheses and theories, to gain generalizable knowledge, and to make it known to the scientific community, e.g., through publications.
- c) It applies the law according to its own interpretation which may differ from the standard interpretations but remains within the scope of possible ones.
- d) All members of the Task Force have a tight time schedule and hence tend to minimize their efforts.

These values are guesses since no investigation was made to assess them. Nevertheless the list seems quite reasonable, although one must assume that this list is not complete and that additional values might be important.

4 KAI KURIE KONFLIKTAI

Konfliktai, kuriuos čia aprašysime, atspindi įvairių sričių įvertintojų nurodytus įvertinimo prieštaravimus (Gastager ir Patry, 2001), kai kurie iš jų jau buvo aptarti teorinėje literatūroje. Darbo grupė tikėjosi susidurti su konfliktais – tai bet kurio įvertinimo projekto atributai, kitaip negalima tikėtis, kad įvertinimas bus sėkmingas. Buvo numatyti labiausiai tikėtinų problemų sprendimo būdai. Bet ne visos problemos galėjo būti numanytos, kai kurie prieštaravimai pasirodė esą sudėtingesni nei buvo tikėtasi.

Pirmasis konfliktas tiesiogiai ar netiesiogiai laikomas visų kitų prieštaravimų, kuriuos čia aptarsime, pagrindu: tai prieštaravimas tarp ekonominių vertybių (5 skiltis; 2 lentelė) ir visų kitų vertybių; šis dalykas nebuvo netikėtas, nes „dėl biudžeto stokos atliekant apklausą ir vertinimo darbą kai kurių gerų pasirinkimų visada reikia atsakyti“ (Cronbach, 1987, p. 322). Pavyzdžiui, geras įvertinimas (2 skiltis) yra brangus; tas prieštaravimas tarp kokybės ir kaštų laikomas kiekybiniu konfliktu priemonių lygmenyje. Tačiau oficialūs pareigūnai (antroji suinteresuotų asmenų grupė) dažnai nepripažįsta įvertinimo procedūrų kaštų ir nesuteikia reikalingos finansinės paramos. Šiuo klausimu plačiau kalbėti nenorime, tačiau reikia pridurti, kad su šiuo prieštaravimu susiduriama nuolat.

4.1. APRĖPTIS PRIEŠ PATIKIMUMĄ

Konfliktas tarp aprėpties ir patikimumo ypač akivaizdus bendravimo srityje (Cronbach, 1987, p. 36f.); šiuo atveju turime galvoje tik siaurą sąvoką, kur teigiama, kad imant tam tikro ilgumo vertinimo priemonę (pvz., anketą) susiduriama su prieštaravimu tarp sričių, kurias galima vertinti, skaičiaus ir vertinimo patikimumo (Cronbach, 1970, p. 179). Tai metodinių vertybių konfliktas (2 skiltis; 2 lentelė), bet jį sukelia poreikis rinkti trumpą priemonę, nes to reikalauja asmeninės vertybės (6 skiltis), kurias gina studentai (4 grupės suinteresuoti asmenys, d vertybė) ir dėstytojai (suinteresuoti 3 grupės asmenys; e vertybė). Tai jau kiekybinis konfliktas remiantis aukščiau pateiktu apibrėžimu – ir priešingai Cronbach terminologijai – tai ne dilema. Šis konfliktas kyla priemonių lygmenyje: iš principo didelio patikimumo tikslas ir daugelis vertinimo sričių yra palyginti suderinami dalykai, tačiau akivaizdu, kad nėra priemonių abu tikslus pasiekti tuo pat metu.

Kaip teigiama Spearman- Brown formulėje (Cronbach, 1970, p. 161 ir 171ff.), šis konfliktas atsiranda dėl to, kad anketos patikimumas priklauso nuo jos ilgio. Dėl šios priežasties anketos, kurios paprastai pateikiamos studentams, kad jie sudarytų dėstytojų reitingus, yra gana ilgos (pvz., Rindermann 2001; 51 klausimas; Marsh, e.g. 1991; 35 klausimai ir t. t.). Tokio ilgumo anketa nebūtų buvusi priimtina nei studentams, nei dėstytojams. Dėl šios priežasties darbo grupė nusprendė pasitelkti kitokio pobūdžio tyrimo metodą, išlaikydama tą patį problemos lygmenį (žr. aukščiau, 1 poskyris), ir pakeisti anketos klausimų struktūrą: vietoj paprastų tradicinių teiginių, pvz., „dėstytojas gerai išaiškina“, darbo grupė pateikė taip vadinamuosius „globalius klausimus“: bendrą dimensijos apibūdinimą su iliustracijomis ir pavyzdžiais skliaustuose mažesnėmis raidėmis. Tokiu atveju visa informacija pateikiama vienu sakiniu, o ne keliais klausimais.

4 SOME CONFLICTS

The conflicts to be discussed below are quite representative for the conflicts in evaluation reported by evaluators in different fields (Gastager & Patry, 2001), and some of them have already been discussed in the literature. The Task Force had anticipated most of them in some way – this is a necessity in any evaluation project, otherwise it is almost certain that it will fail. And it had also conceived some solutions to the most likely problems. But not all problems could be anticipated, and some conflicts turned out quite different than assumed.

A first conflict is underlying, directly or indirectly, all the other conflicts to be discussed below: the conflict between economic values (area 5 in table 2) on one hand and many other values on the other hand; this was no surprise, since “budgets will always force omission of some good choices of directions for inquiry and evaluative activities” (Cronbach, 1987, p. 322). For instance, good evaluation (area 2) is expensive – this antagonism between quality and costs is a quantitative conflict on the level of the means. However, the officials (stakeholder group 2) often do not recognize the costs of the different evaluation procedures and hence do not provide the necessary financial support. We do not want to go into details here, but this issue comes up over and over again.

4.1 BANDWIDTH VS. FIDELITY

The conflict between bandwidth and fidelity has the broad meaning in communication (Cronbach, 1987, pp. 36f.); in the present case, we refer only to the narrow concept according to which for an assessment instrument (e.g., a questionnaire) of a given length there is an antagonism between the number of domains which can be assessed and the reliability of the assessments (Cronbach, 1970, p. 179). This is a conflict within the methodical area (area 2 in table 2) but caused by the need to use short instrument as required by the personal values (area 6) defended by the students (stakeholder group 4, value d) and the teachers (stakeholder group 3, value e). It is a quantitative conflict, hence according to the definition above – and in contrast to Cronbach’s terminology – it is not a dilemma. The conflict is on the level of the means: in principle the goals of high reliability and many domains are compatible, yet it seems that there is no means to achieve both goals simultaneously.

This conflict is due to the fact that the reliability increases with its length according to the Spearman-Brown formula (Cronbach, 1970, pp. 161 and 171ff.). For this reason the typical questionnaires for the student ratings of teaching are quite long (e.g., Rindermann, 2001: 51 items; Marsh, e.g. 1991: 35 items; etc.). A questionnaire of this length, however, would not have been accepted by the students and teachers. For this reason the Task Force decided to use a different approach, leaving the level of the problem (see above, section 1) and changing the structure of the items: Instead of the traditional simple items like “the teacher provides good explanations” the Task Force used so-called “global items”: General descriptions of the dimension and, with smaller charac-

mais. Tipiškas pavyzdys galėtų būti čia pateikiama frazė „*Studentų motyvavimas (skatinimas, susidomėjimo palaikymas ir t. t.)*“, kurios vertinimą reikėjo pasirinkti iš 7 skalėje nuo „labai silpnai“ iki „labai stipriai“. Iš viso buvo pateikta 17 tokių „globalių“ frazių, padedančių įvertinti dėstytojo darbą skirtingose darbo grupės nustatytose srityse; toliau buvo duotas bendro pobūdžio sprendimas, kurio vertinimo skalė nuo „labai gerai“ iki „labai blogai“. Apie tyrimo patikimumą jau kalbėjome (3.1 poskyryje), jo rezultatai pateikti kitose ataskaitose (Lengenfelder ir kt., spaudoje); pasirodė, kad tokios priemonės gana patikimos.

4.2. FORMUOJAMASIS PRIEŠ APIBENDRINANTĮ ĮVERTINIMĄ

Vyriausybė (1 suinteresuota grupė; 3 lentelė), savo reikalavimuose vykdyti įstatymą (c vertybė), atrodo, ypač akcentuoja apibendrinamąjį įvertinimą (a ir b vertybės), o darbo grupė sutelkė dėmesį į formuojamąjį įvertinimą: pareigūnai (2 suinteresuota grupė; žr. a vertybė) turėjo aptarti studentų sudarytus atskirų dėstytojų reitingus su pačiais dėstytojais. Vyriausybė taip pat skatino šią procedūrą (žr. aukščiau, 3.2 poskyris, 1a klausimas), bet tik su dėstytojais, kurių darbas buvo įvertintas blogai; atrodo, kad formuojamasis vertinimas buvo taikomas tik šiems dėstytojams. Darbo grupės nuomone, reitingai turėtų būti aptariami su visais dėstytojais.

Pasirodė, kad ši procedūra sunki ne tik dėstytojams, bet ir asmenims, atsakingiems už grįžtamojo ryšio pateikimą. Akivaizdu, kad nepaisant darbo grupės pastangų informuoti dėstytojus apie formuojamąjį metodą, iš tikrųjų dėstytojai jo nesuprato; be to, buvo laikomasi prielaidos, kad įvertinimas apibendrinamasis; dažnai (pvz. Bauer, 2001) buvo vadovaujamas argumentu, kad jeigu tam tikras dėstytojo kursas taikant tam tikrą skalę buvo labai žemai įvertintas, tai reiškia, kad toks dėstytojas ar dėstytoja, matyt, yra blogi (tai tarsi charakterio savybė ar bruožas, o ne elgsena tam tikroje situacijoje; žr., pvz., Jones ir Nisbet, 1971), todėl visai nestebina tai, kad dėstytojai gynėsi nuo šių insinuacijų. Dėstytojų daromos prielaidos iš tikrųjų nebuvo visiškai klaidingos, nes asmenys, atsakingi už grįžtamąjį ryšį, patys nesuprato šios idėjos ir interpretavo neigiamus vertinimus panašiai kaip ir dėstytojai.

Reikėjo išsamiau informuoti dėstytojus ir asmenis, atsakingus už grįžtamąjį ryšį, surengti susirinkimus ar net organizuoti kursus. Tačiau tai buvo neįmanoma dėl keleto priežasčių. Viena priežastis buvo menkas biudžetas, kita priežastis – giliai įsišaknijęs įsitikinimas, kad žmonės mano žiną, kas yra įvertinimas, nors ir prisipažįsta nesą šios srities specialistai. Mes visi mokėmės ir mokėme kitus, tai reiškia, kad patyrėme kitų įvertinimą ir įvertinome patys. Vadinas, net jei jie ir supranta, kad nėra ekspertai, vis tiek yra įsitikinę, kad išmano kas yra švietimas ir įvertinimas, o šiuos jų įsitikinimus pakeisti sunku. Jei dėstytojams ir asmenims, atsakingiems už grįžtamąjį ryšį, pateikta informacija apie įvertinimą neatitinka jų nuomonės, galimas daiktas, kad tas skirtumas bus ignoruojamas, o žmogus ir toliau laikysis savo buvusių įsitikinimų, o gal šis asmuo iš viso nevaikšto į tokius informacinio pobūdžio susirinkimus, nes nenori leisti laiko tam, ką ir taip mano jau žinąs.

ters in brackets, some illustrations or examples. Instead of using several items, the full information is put into one item. A typical item is “Motivating the students (stimulation, furthering interest, etc.)”, to be rated on a 7-point scale from “very weakly” to “very strongly”. Overall, 17 such global items were used to cover the different domains the Task Force had conceived; further, a general judgment was included, with the ratings ranging from “very bad” to “very good”. The validation study was already mentioned above (section 3.1), its results are reported elsewhere (Lengenfelder et al., in press); the instrument proved to be quite satisfactory.

4.2 FORMATIVE VS. SUMMATIVE EVALUATION

The government (stakeholder group 1, table 3), in its enforcement of the law (value c), seems to put particular emphasis on a summative evaluation (values a and b), while the focus of the Task Force was on a formative evaluation: The officials (stakeholder group 2; see value a) were supposed to discuss the results of the ratings with the individual teachers. This procedure was also stipulated by the government (see above, section 3.2, issue 1a), but only for teachers with low ratings; at least for them a formative concept seems to have been aimed at. According to the Task Force this should be done for all teachers.

As it turned out, this kind of procedure is difficult to convey both to the teachers and to the officials who were supposed to give the feedback. Obviously, despite the efforts of the Task Force to inform about the formative approach the teachers did not really understand it; instead, the underlying assumption was that the evaluation was summative; frequently (e.g., Bauer, 2001) it was argued that if a teacher had a low rating in a particular course with a particular scale, this meant that he or she was supposedly a bad teacher (attribution as disposition or trait instead of behavior in a particular situation; see, for instance, Jones & Nisbet, 1971), and not surprisingly the teachers defended themselves against this insinuation. Actually the assumption of the teachers were not completely false because the officials interpreted the negative ratings similarly because they did not understand the idea either.

It would have been required to inform the teachers and officials more deeply in meetings or even courses. However, this was not possible for several reasons. The low budget was one reason, but another is rooted in the fact that people believe they know what evaluation is although they acknowledge that they are not specialists. We all have experienced education and have practiced education; this means that we have experienced and practiced evaluation. In consequence, even if they know that they are not experts, they have strong beliefs about education and evaluation which are hard to replace. If the information about evaluation provided to the teachers and officials does not comply with the presumed knowledge, the difference is likely to be neglected and the person sticks to his or her initial beliefs – or maybe he or she does not go to such information meetings at all because he or she does not want to spend time for something he or she thinks to know already.

There are several conflicts. A first one is between summative and formative evaluations – this is on the goal

Šiuo atveju susiduriame su keliais prieštaravimais. Pirmasis susidaro tarp apibendrinamojo ir formuojamojo įvertinimo – tikslų lygmenyje, kiekybinis vertinimas (formuojamas ar apibendrinamasis, bet galime apmąstyti ir rasti būdų jungti tuos du metodus). Antrasis susidaro tarp reikalingos informacijos perdavimo ir turimų resursų (kiekybinis; priemonės). Trečiasis – tarp idėjų, kurių laikėsi darbo grupė, ir idėjų, kurių laikėsi dėstytojai ir pareigūnai. Prieštaravimų priežastis nėra anksčiau minėta struktūrinė klasifikacija, tai greičiau įgyvendinimo problema, su kuria susiduriame daugelyje sričių.

4.3. REGLAMENTUOTOS NORMOS

Vienas klausimas buvo keliamas nuolat: ką reiškia geras dėstymas? Ar vertinimas ketvertu – teoriškai vidutinis – rodo gerą dėstymą, ar tik vidutinišką ar net blogą dėstymą? Sprendimas apie atskiro objekto vertę priklauso nuo to, su kuo vertinamas objektas yra lyginamas. Paprastai išskiriami trys nustatytų normų tipai (ypatingais atvejais galima pridėti ir daugiau tipų, bet šiuo atveju pakanka trijų): individualios, socialinės ir į kriterijus orientuotos vertinimo normos. Individualiai nustatytos normos reiškia, kad individo (šiuo atveju dėstytojo) ankstesnis darbas (šiuo atveju vieno ar kelių ankstesnių reitingų rezultatai) yra lyginami su dabartiniu darbu. Kadangi įvertinimas nėra laikomas ilgalaikiu nuosekliu tyrimu, tai neįmanoma kontroliuoti visų kintamųjų, kurie gali turėti įtakos (pvz., tas pats dėstytojas turėtų dėstyti tą patį kursą panašioms studentams panašiomis sąlygomis), todėl šis įvertinimas negali remtis individualiomis normomis. Taigi arba socialinės, arba į kriterijus orientuotos normos yra galimos. Pirmoji reiškia, kad kažkas yra vertinamas gerai, jei jis yra geresnis nei jo kolegos, o į kriterijus orientuotas įvertinimas reiškia, kad individo darbas vertinamas pagal tam tikros instancijos sukurtus kriterijus, ir su šiais kriterijais individo darbas yra lyginamas.

Įstatymuose (4 vertybių skiltis, 2 lentelėje) nėra aiškiai nurodyta, kokios nustatytos normos turi būti taikomos. Kaip minėta aukščiau (4.2 poskyris), asmenys, atsakingi už grįžtamąjį ryšį, turi su neigiamus vertinimus gavusiais dėstytojais aptarti galimą jų darbo pagerinimą. Šį nurodymą išleidusi ministerija atsisakė nuostatos remtis socialine norma, kur reikalaujama diskusijos tuo atveju, kai reitingai yra daug žemesni už vidutinius tam tikro universiteto dėstytojų reitingus. Tad galima daryti prielaidą, kad remiamasi į kriterijus orientuota norma.

Formuojamojo įvertinimo požiūriu (žr. aukščiau; 4.2 poskyris) į kriterijus orientuotų normų taikymas tinkamas, jei tik galima juo remtis. Tačiau būtina tokius kriterijus apibrėžti, o tai jau yra institucijos ar jos atstovų užduotis (2 suinteresuotų asmenų grupė); pavyzdžiui, universitetas galėtų teigti, kad ypač vertinami puikūs studentų pasiekimai ar gera dėstymo atmosfera arba studentų kritinio mąstymo ugdymas ir t.t. Galima daryti prielaidą, kad dauguma dėstytojų šių skirtingų normų pasiekimą taip pat laiko prieštaringu, kaip ir kiekybinius priemonių konfliktus: tikslai nebūtinai nesuderinami, bet dauguma dėstytojų mano, kad vieno iš tikslų siekimas gali sukliudyti siekti kitų tikslų.

Kokį kriterijų pasirinkti – tai strateginis universiteto sprendimas. Darbo grupė nesijautė esanti kompetentin-

level and quantitative (formative or summative, but one can imagine some ways to combine the two approaches). A second is between the necessary information transmission and the available resources (quantitative; means). A third one is between the ideas defended by the Task Force and those defended by the officials and teachers. This is not a conflict in the sense of the above taxonomy; rather it is an implementation problem which is encountered frequently in domains.

4.3 THE REFERENCE NORMS

One question was asked again and again: What is good? Is a rating of “4” – the theoretical average – a sign of good teaching, or does it indicate only a medium performance, or maybe even a bad one? The judgment of a single value depends on the reference to which it is compared. Typically one distinguishes three types of references (for specific cases one can add more types, but here these are sufficient): individual, social, and criterion oriented reference norms. An individual reference norm means that an individual’s (here: a teacher’s) previous performance (here: ratings of one or more earlier assessments) are compared to the present performance. Since the evaluation is not conceived as time series study and it is impossible to control for all the variables that may have an influence (e.g., the same teacher should teach the same course with similar students under similar conditions), the focus of the present evaluation study cannot be on the individual norms. Hence, either the social or the criterion oriented norm is possible. The former one states in that someone is judged to be good if he or she is better than his or her peers, whereas the latter refers to a criterion defined by some instance and with which the performance of the individual is compared.

In the law (values area 4 in table 2) it is not clearly stipulated what type of reference norms are to be used. It was mentioned above (section 4.2) that the officials have to discuss possible improvements with the teachers who have clearly negative ratings. The ministry which issued this rule resisted to the temptation to refer to the social norm, such as requiring discussions when the ratings are clearly below the average ratings of the teachers of the respective University. Hence one could assume that the underlying reference norm is criterion oriented.

From the point of view of a formative evaluation (see above, section 4.2) the use of criterion oriented norms – if available – is appropriate. This requires, however, that such criteria are defined, and this is the task of the institution or its representatives (stakeholder group 2); for instance, the University could say that it puts particular emphasis on high achievements of the students, or on a good teaching climate, or on educating the students to critical thinking, etc. One can assume that most teachers regard the attainment of these different norms as antagonistic, as quantitative conflict of the means: the goals are not necessarily mutually exclusive, but most teachers think that aiming at one of the goals will inhibit aiming at the other ones.

It is a strategic decision of the University what criterion to prefer. The Task Force did not feel competent to make such a decision, rather it suggested to the rector and vice-rectors to do so, but so far there has not been

ga priimti tokį sprendimą, dėl to buvo pasiūlyta jį priimti rektoriui ir prorektoriams, bet kol kas jokių sprendimų šiuo klausimu nepriimta. Žinoma, tai priešprieša įvertinimo sąvokai, bet asmenims, atsakingiems už grįžtamąjį ryšį, to negalima perteikti (žr. 2 grupė, a tikslas; 3 lentelė: ekspertizės stoka).

Socialinių vertinimo normų taikymas formuojamojo įvertinimo tikslams neturi prasmės, nes įvertinimo tikslas yra pateikti atskirą grįžtamąjį ryšį atskiriems dėstytojams bei detalį stiprių ir silpnų pusių analizę bei pasiūlyti galimybes tobulėti. Tinkamesnis į kriterijus orientuotas įvertinimas, nors universaliųjų normatyvų neturėtume taikyti, nes atitinkamomis sąlygomis reikalaujama taikyti skirtingus kriterijus. Pavyzdžiui, net jei visame universitete keliamas tikslas pasiekti puikių laimėjimų, vis tiek bus tokių kursų, kaip pavyzdžiui, komunikacijos ir socialinių kompetencijų įgijimo problemoms skirti kursai, kurių pagrindinis tikslas ne geri pasiekimai, o socialinis mokymasis, kuriam reikalinga gera socialinė atmosfera.

Kitas metodas – į individualius kriterijus orientuotų normų taikymas, t.y. dėstytojo klausima, kaip jis norėtų būti vertinamas ar į kokius kriterijus, jo nuomone, vertintojai turėtų labiau kreipti dėmesį (dėl skirtingų tikslų sukeltų prieštaravimų žr. aukščiau). Mes šį metodą taikėme kitame kontekste. Tačiau klausiti dėstytojų, kokiems kriterijams jie teiktų pirmenybę, rizikinga, nes tai prieštarautų dėstytojų tikslui minimalizuoti pastangas (3 grupė, e tikslas) ir galėtų sukelti pasipriešinimą, jei šis metodas bus taikomas dideliu mastu.

Kadangi nebuvo galima taikyti jokių į kriterijus orientuotų normų – nei bendrųjų, nei individualių, tai darbo grupė turėjo pasitelkti primityvias socialinių vertinimų normas (žr. pavyzdžius apie grįžtamąjį ryšį www.sbg.ac.at/evaluation): dėstytojas buvo supažindinamas su gautais vertinimo rezultatais kiekvienu klausimu (kadangi pasiskirstymas buvo labai asimetriškas, tai nuspręsta apskaičiuoti medianą, nes vidurkis galėjo būti klaidingas). Toliau buvo nurodytos 20-oji ir 80-oji procentilės. Pagal apibrėžimą 20 proc. viso universiteto studentų (arba pagal tokią proporciją parinkti iš viso įvertinimui atrinktų dalyvių skaičiaus) pateikia žemesnį vertinimą nei 20-oji procentilė, o kiti 20 proc. – viršija 80-ąją procentilę atitinkamoje skalėje; 60 proc. vertinimas pasiskirsto tarp 20-osios ir 80-osios procentilės. Lyginimas su visais universiteto studentais, neatsižvelgiant į tai, koks jų skaičius būtų kiekviename kurse, koks kurso turinys, tipas ir t.t., turi būti išdiskutuotas. Reikia pripažinti, kad statistiniu požiūriu (2 vertinimo sritis, a pavyzdys, 2 lentelė), šis pavyzdys netinkamas (žr. Bauer, 2001); pavyzdžiui, skirtingų kursų vertinimui reikėtų taikyti skirtingas socialines normas. Tačiau čia iškyla keli prieštaravimai (priemonių, o ne tikslų lygmenyje), į kuriuos reikia atsižvelgti:

- Visų pirma, skirtingų socialinių normų taikymas reikalauja papildomos informacijos apie kursus. Bauer užsimena apie kursų lankančių studentų skaičių; galima pridėti ir kitas kurso charakteristikas, pvz., turinį (pvz., neteisinga lyginti statistikos kursą su socialinių kompetencijų lavinimo užsiėmimais, nes pirmasis yra daug sunkesnis ir ne taip studentų mėgstamas), bet darbo grupė neturėjo visų svarbiausių parametrų (2 kokybinis konfliktas vertybių srityje).

a decision in this regard. Of course this is opposed to the concept of evaluation, but this could not be conveyed to the officials (see group 2, aim a in table 3: lack of expertise).

The use of social reference norms for formative purposes does not make sense since the aim of the evaluation is to provide an individual feedback to the individual teacher combined with a detailed analysis of the strengths and weaknesses and to suggest possibilities of improvement. The criterion oriented norm is much more appropriate, although maybe one should not use a universal norm since the respective conditions may require different criteria. For instance, even if the whole University is supposed to aim at high achievements, there are some courses like communication and social competence trainings in which instead the primary aim is social learning requiring a good social atmosphere.

Another approach could be to use an individual criterion oriented norm, namely by asking the teacher how he or she expects to be rated or what criterion he or she wants to pay particular attention to (with the conflict between different aims, see above, in mind). We have done this in another context; asking the teachers what criterion they prefer, however, it is against their goal to minimize effort (group 3, aim e) and will encounter opposition if done on a large scale.

Since no criterion oriented norm was available – neither general nor individual norm –, the Task Force had to recur to the admittedly primitive social reference norm (examples of the feedback are available in www.sbg.ac.at/evaluation): The teacher's average rating for each item on his or her lecture (more precisely: the median, which was chosen because the distributions were skewed so the average could be misleading) was reported to him or her. Further, the 20th and 80th percentiles were indicated. By definition, 20% of the students of the whole University (or if so chosen of a sub-sample) give a rating below the 20th percentile and 20% of them give a rating above the 80th percentile on the particular scale; 60% have a rating between the 20th and the 80th percentile. The comparison with all students of the University, irrespective of the number of students in the course, the content, the type, etc. needs to be argued for. It must be acknowledged that from a statistic point of view (values area 2, example a in table 2), this approach is inappropriate (see Bauer, 2001); for instance, it would be appropriate to use different social norms for different types of courses. However, there are several antagonisms (on the level of the means, not of the goals) which must be taken into account:

- First of all, the use of different social norms would require additional information about the courses. Bauer mentions the number of students participating in the course; one could add other characteristics of the courses, such as the content (e.g., it is unfair to compare courses in statistics with social competence trainings because the former are much more difficult and less appreciated by the students) – but the Task Force did not know all the relevant parameters (qualitative conflict within values area 2).

- Second, the explanation of the statistical procedure to the teachers would have been even more com-

- Antra, statistinių procedūrų aiškinimas dėstytojams galėjo būti dar sudėtingesnis nei ši procedūra – daugeliui dėstytojų parinkti metodai ir taip kėlė problemų (kokybinis konfliktas 3 suinteresuotų asmenų grupėje, d vertybės, 3 lentelė).

- Trečia, tokiai analizei pateikti trūko ekonominių resursų (5 vertybių skiltis), ir nebuvo galima akcentuoti kitų tikslų, nors jie buvo laikomi dar svarbesniais (kiekybinis konfliktas).

Nevisiškai tinkamų statistinių metodų naudojimas gali būti pateisinamas žvelgiant iš etinių pozicijų (1 vertybių skiltis), nes reglamentuotos normos nėra tokios svarbios formuojamajam įvertinimui, jos svarbesnės apibendrinančiam įvertinimui; reglamentuotos normos buvo pateiktos informaciniais tikslais ir neakcentuojant minties, kad žemesnį nei 20-oji procentilė įvertinimą gavę dėstytojai buvo „blogi“ dėstytojai, o įvertinti aukščiau „vidutiniai“ ar „geri“ (nors dėstytojai įvertinimus suvokė būtent taip). 20-osios procentilės pasirinkimas yra tikrai sutartinis (taip buvo pasakyta ir dėstytojams) ir neturėtų būti suvokiamas kaip kritinė riba.

4.4. ĮVERTINIMO PASEKMĖS

Įvertinimai prasmingi tik tada, jei vertinimai turi pasekmes. Formuojamojo vertinimo koncepcijoje vertinimo pasekmė yra tiesioginis grįžtamasis ryšys, kuris sąlygoja praktinę kaitą, nes dėstytojas, išklausęs vertinimo rezultatus, daro išvadas ir atitinkamai keičia savo tiesioginę veiklą. Bet tokius pokyčius buvo galima priversti daryti tik tam tikru laipsniu (žr. 4.2 ir 4.3 poskyriai), nes nei universiteto administracija, nei dėstytojai iki galo nesuprato šios koncepcijos.

Vietoj formuojamojo atliekant apibendrinamąjį įvertinimą, pasekmių sulaukiama kitame lygmenyje: už gerą dėstymą galima apdovanoti, o už blogą dėstymą „nubausti“. Kita vertus, vertinimo rezultatais galima remtis priimant į darbą, aukštinant pareigose ir atliekant kitus su karjera susijusius sprendimus (žr. 1 suinteresuotųjų grupė, a tikslas, 3 lentelė).

Galimų veiksmų pasirinkimas neigiamai įvertintų dėstytojų atžvilgiu („nuobaudos“) gana negausus. Dėstytojo iš darbo atleisti negalima, už, kaip manoma, blogą dėstymą dėl teisinių (4 vertės skiltis, 2 lentelė) ir asmeninių priežasčių (6 vertės skiltis), lieka tik keletas kitų priemonių (ir aukščiau minėti formuojamojo vertinimo metodai). Socialiniu požiūriu (3 vertės sritis), be to, sutarus, ką galima vadinti „geru dėstymu“ (2 vertės skiltis), reikėtų akcentuoti tik apdovanojimus už gerą dėstymą, o ne nuobaudas už prastą dėstymą. Tačiau universiteto pareigūnai nepalaikė darbo grupės pasiūlymų šiuo klausimu, jiems pasirodė svarbiau nubausti dėstytojus už „blogą“ dėstymą.

Šiame kontekste reikia paminėti pakartotinus universiteto valdžios prašymus paskelbti atskirų įvertinimų rezultatus. Darbo grupė atvirai pasisakė prieš tokius siūlymus, nes jie prieštaravo formuojamojo įvertinimo tikslams. Šiuo požiūriu įstatymas neiškus ir gali būti interpretuojamas įvairiai. Tačiau nors darbo grupė ir prieštaravo, atskiri įvertinimo rezultatai buvo paskelbti vieno fakulteto internetiniame puslapyje, kituose fakultetuose rezultatus, nors ir sunkiai, bet buvo galima gauti. Tai jau kokybinis konfliktas, arba duomenų prieinamumo atžvilgiu, kieky-

plicated than with the current procedure – and many teachers had their problems already with the methods chosen (qualitative conflict with stakeholder group 3, values d in table 3).

- Third, the economic resources (values area 5) were not sufficient to provide such analyses, or other aims which were judged more important would have had to be reduced (quantitative conflict).

The use of statistical approaches which were not absolutely appropriate is ethically justifiable (values area 1) because the reference norm has not such an importance in formative evaluations as it would have in summative evaluations; the reference norms were provided for information purposes and not with the idea that those teachers below the 20th percentile were “bad” teachers and those above it “medium” or “good” ones (although it was perceived so by the teachers). The choice of the 20th percentile was absolutely arbitrary (and this was also communicated to the teachers) and must not be taken as a critical mark.

4.4 CONSEQUENCES OF THE EVALUATION

Evaluations make sense only if the assessments have consequences. In a formative concept the consequences are in the direct feedback which supposedly yields changes in practice since the teacher draws the conclusions from the assessments him- or herself and acts consequently. But this could be enforced only to a little degree (see sections 4.2 and 4.3) because neither the officials nor the teachers really understood the concept.

Assuming a summative instead of a formative evaluation, the consequences can be sought on a different level: one can reward good teaching and “punish” bad teaching. On the other hand, the results of the ratings can be used for hiring, promotion, and other career decisions for the individual teacher (see stakeholder group 1, aim a in table 3).

The scope of possible actions with respect to teachers with negative ratings (“punishment”) is very small. One cannot dismiss teachers because of supposedly bad teaching for legal (values area 4 in table 2) and personal reasons (values area 6), and very few other means (besides the formative approaches mentioned above) are available. But from the social point of view (values area 3), as well as in agreement with what can be called “good evaluation” (area 2), the focus should be laid on rewarding good teaching instead of punishing bad one. However, the propositions of the Task Force in this regard were not accepted by the officials of the University, and for them the focus remained on punishing “bad” teachers.

In this context one must also refer to the repeated request by officials to publish individual evaluation results. The Task Force objected explicitly against such propositions because they are in contrast to a formative endeavor. In this regard, the law is not clear – it can be interpreted in different ways. Against the Task Force’s appeal the individual results were published – in one Faculty in the internet, the other ones, the results were available, yet difficult to get. This is a qualitative conflict – or with regard to the degree of accessibility of the data a

binis konfliktas: tikslų lygmenyje (viešas rezultatų skelbimas) asmeninių vertybių (6a, 2 lentelė) ir to, kaip pareigūnai įstatymą interpretavo (4 skiltis, 2 lentelė; 2 suinteresuotų asmenų grupės, 3 lentelė).

4.5. STUDENTŲ REITINGAI IR ĮVERTINIMAS

Įstatyme ypač pabrėžiami studentų sudaryti dėstytojų reitingai. Diskusijos su dėstytojais ir asmenimis, atsakingais už grįžtamąjį ryšį, parodė, kad tik studentų reitingai buvo laikomi svarbiu dalyku, nepaisant darbo grupės pastangų pabrėžti tai, kad įvertinimui taikant tik dėstytojų reitingavimą praleidžiamos kelios labai svarbios sritys. Tiesą sakant tai natūralistinis sofizmas, klaidinga argumentacija (išvada apie tai, kaip kas nors egzistuoja, pateikta taip, tarsi iš tikrųjų taip turėtų būti), todėl, kad negalima sakyti, kad jei studentai mano, jog tam tikras dėstytojų būdas yra geras, jis iš tikrųjų geras; pavyzdžiui, daugelis dėstytojų įsitikinę, kad studentai negali spręsti apie dėstytojų kompetentingumą. Todėl buvo svarbu rasti nepriklausomus kriterijus (žr. 4.3 poskyrį), norint nustatyti, ar dėstytojas geras, ar ne.

Neabejojame studentų pateiktos informacijos patikimumu. Teorinėje literatūroje aprašyti įvairūs efektai, pavyzdžiui, Dr. Fox-efektas (Naftulin, Ware ir Donnelly cituota iš Marsh ir Ware, 1982: kai aktorius labai išraiškingai dėstė nesąmoningą turinį ir susilaukė aukštų įvertinimų), pažymių rašymo atlaidumas (Greenwald ir Gillmore, 1997: studentai vertina geriau, jei tikisi gauti geresnius pažymius), lyčių efektas (reitingai priklauso nuo to, ar dėstytojas ir studentai tokios pat lyties ar ne). Įvairių studijuojamų dalykų rezultatai sudėtingi, pavyzdžiui, lyčių efektas atveju: kai kuriuose tyrimuose (pvz., Tatro, 1995), vaikinai studentai geriau įvertino dėstytojas, o merginos studentės dėstytojus; kituose (aprašyta Smith ir kt., 1994) – vaikinai studentai geriau įvertino dėstytojus, o merginos studentės – dėstytojas. Apskritai šie efektai statistiškai reikšmingi, jei tyrimo dalyvių skaičius palyginti gausus, bet efekto dydis nežymus ir gali būti ignoruojamas dėl praktinių sumetimų (Gigliotti ir Buchtel, 1990) ir nekelti jokios grėsmės patikimumui, jei vertinimas pagrįstas logišku teoriniu pagrindu.

Tačiau dėstytojas apima ne vien tai, ką gali reitinguoti studentai. Pavyzdžiui, klausimai „Ar tinkama egzaminų sistema?“, „Ar studentai išmoko tai, ką turėjo išmokti?“, „Ar tinkamas dėstytojų turinys?“ ir t.t. yra svarbūs, bet j juos negalima atsakyti remiantis dėstytojų reitingais. Tai konfliktas tikslų lygmenyje: kokio pobūdžio informacijos klausiami? Šiuo atveju viena vertus yra įstatyminės vertybės, nes įstatymuose akcentuojamas studentų reitinguojamas dėstytojas, kita vertus profesionalumo atliekant įvertinimą kriterijai (2f vertė, 2 lentelė) ir kokybės vadybos vertybės, kurias turėtų ginti asmenys, atsakingi už grįžtamąjį ryšį (2 suinteresuotų asmenų grupė), bet jie to nedaro, nes nėra šios srities specialistai.

4.6. DĖSTYTOJŲ ĮTAKA TYRIMO REZULTATAMS

Per daug neteikdami tam dėmesio, norėtume paminėti ir tai, kad kai kurie dėstytojai bandė daryti didesnę ar mažesnę įtaką studentų atsakymams. Neoficialiai minimi tokie dėstytojų poveikio būdai:

quantitative one – on the level of the aims (public availability of the results) between the personal values (6a in table 2) and the officials' interpretation of the law (area 4 in table 2; stakeholder group 2 in table 3).

4.5 STUDENT RATINGS VS. EVALUATION

The law puts a strong emphasis on the student ratings of teaching. Similarly, the deliberations with the teachers and officials showed that only the student ratings were perceived as relevant, despite the efforts of the Task Force to stress that several important areas of teaching are left out if only student ratings are used in the evaluation. Actually it is a naturalistic fallacy (a conclusion from how something *is* to how something *should be*) to say that since the students think a certain teaching is good it is indeed good; for instance, many teachers objected that students were not able to judge whether the teacher was competent or not. It would be important to have independent criteria (see section 4.3) to say whether a teaching is good or not.

We do not doubt the validity of the information by the students. Several effects have been discussed in the literature such as the Dr. Fox-effect (Naftulin, Ware & Donnelly, quoted from Marsh & Ware, 1982: An actor teaches very eloquently nonsense content and gets high ratings), grading leniency (Greenwald & Gillmore, 1997: The ratings are better if good grades are expected), gender effects (the ratings depend on whether the teacher and the students are of the same sex or not). The results of the different studies are complex, as can be seen with regard to the gender effects: In some studies (e.g., Tatro, 1995), male students give better ratings to female teachers and female students to male teachers; in others (reported by Smith et al., 1994), male students give better ratings to male teachers and female students to female teachers. In general, these effects are statistically significant given the large samples which are available, but the effect sizes are rather small and can be neglected for practical purposes (Gigliotti & Buchtel, 1990) and no real threat to validity if the assessment is based on a sound theoretical framework.

There is more in teaching, however, than what is rated by the students. Questions like “Is the examination system appropriate?”, “Did the students learn what they were supposed to learn?”, “Is the teaching content appropriate?”, etc., are important but cannot be answered with student ratings. This is a conflict on the level of the aims: what kind of information is asked for? In the present case, there are, on one side, legal values because the focus of the law is on student rating; on the other side there are values of professionalism in evaluation (value 2f in table 2) and of quality management which should be defended by the officials (stakeholder group 2) but actually are not because they are not specialists in the domain.

4.6 CHEATING TEACHERS

Without attributing too much importance to it, we want to mention the fact that some teachers tried to influence more or less directly the responses of their students. Several ways to do so were reported informally:

- The teacher watches the students responding to the questionnaire.

- Dėstytoja(s) stebi, kaip studentai atsakinėja į anketos klausimus.
- Dėstytoja(s) pats sako studentams, kaip atsakyti.
- Dėstytoja(s) prašo studentų teigiamai įvertinti, sakydami, kad jų karjera priklauso nuo šio įvertinimo (nors taip nėra).

Galima kelti prielaidą, kad tai kokybinis konfliktas tarp sąžiningumo (1 skiltis, a pavyzdys, 2 lentelė) ir dėstytojo asmeninių interesų (asmeninės vertybės, 6 skiltis), atsižvelgiant į tikslus (šališki prieš nešališkus reitingus; taip pat žr. 3a; 3 lentelėje). Darbo grupė dėjo daug pastangų, kad būtų išvengta tokio apgaulingo elgesio, bet buvo neįmanoma pasiekti, kad vertinimo procesą auditorijose organizuotų pašaliniai žmonės; be to, darbo grupė tikėjo dėstytojų sąžiningumu ir darė prielaidą, kad jie sudarys sąlygas įvertinti teisingai.

Smulkiau šios problemos nenorime aptarti. Ypač dėl to, kad nėra žinoma, ar dažnai kartojosi toks apgaulingas elgesys, nes darbo grupė apie tai pasiekdavo tik nesisteminę informaciją. Reikia pripažinti, kad vis dėlto rezultatus bandoma paveikti, o toks elgesys kelia rimtą pavojų įvertinimo rezultatams. Tačiau darbo grupė neturėjo priemonių, kaip tai apriboti. Akivaizdu, kad *a priori* negalima spręsti apie įvertinamų žmonių sąžiningumą. Galbūt reiktų prašyti studentus protokoluoti vertinimo procesą, kad būtų išvengta tokių manipuliacijų, bet tai sukeltų papildomų konfliktų, susijusių su studentų ir dėstytojų pasitikėjimu (1 sritis) ar nepasitikėjimo skatinimu (3 sritis), ir reikalautų didesnių studentų pastangų (4 suinteresuotųjų asmenų grupė), dėl to išaugtų vertinimo kaštai ir t.t.

5 IŠVADOS

Buvo pateikta įvertinimo prieštaravimų struktūrinė klasifikacija, kurią sudaro (1) bendroji konfliktų rūšių tipologija (1 lentelė), (2) bendrasis vertybių sričių sąrašas (2 lentelė), ir (3) suinteresuotųjų asmenų sąrašas (šis sąrašas panaudotas kaip pavyzdys; žr. 3 lentelę) ir jų vertybės, kurios negali būti taikomos bendro pobūdžio įvertinimui. Bendrosios vertybių sritys ir specifinės suinteresuotųjų asmenų vertybių sritys smarkiai kertasi, todėl šiame straipsnyje vienur kalbėjome apie pirmąsias, kitur apie antrąsias.

Pasirodo, atliekant didelės apimties įvertinimą, tokį kaip ir mūsų pateiktas, kyla daug antagonistinių situacijų. 4 dalyje duotas prieštaravimų sąrašas nėra išbaigtas, galima pridėti dar keletą konfliktų, bet daugelis jų nėra labai svarbūs. Čia pateikta struktūrinė klasifikacija ir metodai atrodo tinkami tokių konfliktų analizei; tačiau ši analizė buvo atlikta *ex post facto*, nors kai kurie potencialūs prieštaravimai buvo numatyti (žr. darbo grupės *Evaluation von Lehrveranstaltungen an der Universität Salzburg*, 2002), pavyzdžiui, aprėpties ir patikimumo dilema, kuriai spręsti ieškota specifinio būdo ir toks būdas išbandytas. Buvo numatyti ir ekonominių vertybių konfliktai, tačiau vis dėl to jie liko milžiniški viso vertinimo proceso metu. Kiti prieštaravimai, pvz., apgaulingas dėstytojų elgesys darbo grupei buvo netikėtas. Siūloma struktūrinė klasifikacija remtis tada, kai įvertinimas suvokiamas kaip būdas imtis priemonių išvengti ar sumažinti galimus konfliktus dar prieš pradėdant įvertinti.

- The teacher tells the students directly what to answer.
- The teacher asks the students to give positive ratings since his or her career would depend on the evaluation (which actually was not the case).

One can assume that this is a qualitative conflict between honesty (area 1, example a in table 2) and personal interests of the teachers (personal values, area 6) with respect to the aims (biased vs. unbiased ratings; see also 3a in table 3). The Task Force had put much effort in preventing such cheating behavior, but it was impossible to have external people going into the classes for the assessment; also the Task Force had assumed that the teachers would provide fair assessment conditions.

We do not want to discuss this problem into more detail. Particularly, it is not known how frequent such a cheating was since the Task Force was informed unsystematically that such behavior occurred. It must be acknowledged that cheating happens and that this is a serious threat to the evaluation. But the Task Force had no control over such behavior. Obviously fairness of the evaluated people cannot be assumed a priori. Maybe one should ask the students to write protocols about the assessment to avoid such manipulations, but this would create additional conflicts, involving issues like trust (area 1) or communicating distrust to both the students and the teachers (area 3), increasing the required student efforts (stakeholder group 4) and the cost of the evaluation, and so on.

5 CONCLUSIONS

The taxonomy of antagonisms in evaluation was presented, consisting of (1) a general typology of conflict types (table 1), (2) a general list of values areas (table 2), and (3) a list of stakeholders and their values specific to the evaluation used as example (table 3) and which hence cannot be proposed for evaluations in general. There is a large overlap between the general values areas and the specific stakeholder values, so in the presentation we referred sometimes to the former, sometimes to the latter.

It turned out that there are many antagonistic situations in a large evaluation like the one presented here. The list of antagonisms presented in section 4 is not complete at all, one could add several conflicts, but most of them are of less importance. The taxonomy and the approach presented here seemed quite useful for such an analysis of conflicts; however, here this analysis was made *ex post facto*, although some of the potential antagonisms were anticipated (see Task Force "Evaluation von Lehrveranstaltungen an der Universität Salzburg", 2002), such as the bandwidth-fidelity dilemma for which a specific solution was sought and tested. The conflicts with economic values were also anticipated, yet remained overwhelming all over the evaluation process. Other antagonisms took the Task Force by surprise, such as the cheating behavior of teachers. It is suggested that the taxonomy should be used when an evaluation is conceived to take measures to avoid or minimize potential conflicts before the evaluation starts.

LITERATŪRA / REFERENCES

- (AERA) (1992). Ethical standards of the American Educational Research Association. URL: <http://www.aera.net/resource/ethics.html> (November 2, 2003).
- Bauer W. (2001). Stellungnahme zur Auswertung der Evaluation der Lehrveranstaltungen an der Universität Salzburg SS 2001 [Statement about the results of the course evaluation at the University of Salzburg in 2001]. Universität Salzburg. URL: <http://www.sbg.ac.at/evaluation/bauerevaluationss01.doc> (18. Juni 2003)
- Baumann U., Bucher A., Ehmer J., Haider G., Paschon A., Patry J.-L. (1999). Bericht zur Beurteilung der Lehrveranstaltungen durch die Studierenden an der Universität Salzburg [Report on the student ratings of teaching at the University of Salzburg]. Universität Salzburg. URL: <http://www.sbg.ac.at/evaluation/ag-berichtnov1999.html> (30. Juni 2003).
- Bundesministerium für Wissenschaft und Verkehr (1997). Verordnung des Bundesministers für Wissenschaft und Verkehr über Grundsätze für die Durchführung von Evaluierung in Forschung und Lehre der Universitäten (Evaluierungsverordnung - EvalVO) [Rules prescribed by the Austrian minister for science and traffic about the principles for the evaluation in science and teaching at the universities]. BGBl. Nr. II 224/1997.
- Campbell D.T. (1969). Prospective: Artifact and control. In: Rosenthal R., Rosnow R. (Eds.): *Artifact in behavior research*. New York: Academic Press, p. 351-382.
- Campbell D.T., Fiske D.W. (1959): Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, p. 56, p. 81-105.
- Cronbach L.J. (1970). *Essentials of psychological testing*. New York: Harper and Row (third edition).
- Cronbach L.J. (1985). *Toward reform of program evaluation*. San Francisco: Jossey-Bass (third edition).
- Cronbach L.J. (1987). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Deutsche Gesellschaft für Evaluation (2002). *Standards für Evaluation (Standards of the German society for evaluation)*. Köln: Eigenverlag.
- Feather N.T. (Ed.) (1982). *Expectations and actions: Expectancy-value models in psychology*. Hillsdale N.J.: Erlbaum.
- Gastager A. (1999). The wider the choice, the greater the responsibility. Balancing between selfishness and altruism in daily school life. In: Patry J.-L., Lehtovaara, J. (eds.). *European perspectives on teacher ethics*. Tampere (Finland): Reports from the Department of Teacher Education in Tampere University, A20, p. 63-74.
- Gastager A., Patry J.-L. (2001). Dilemmastruktur von Evaluationen – Wertpluralismus in Evaluationsprozessen im Bildungsbereich [The dilemma structure of evaluation – values pluralism in educational evaluation processes]. *Internal report to the Austrian – FSR/ project: evaluation in education*, P13594-SOZ.
- Gigliotti R.J., Buchtel F.S. (1990). *Attributional bias and course evaluations*. *Journal of Educational Psychology*, 82, p. 341-351.
- Greenwald A.G., Gillmore G.M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, p. 1209-1217.
- Jones E.E., Nisbett R.E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. In: Jones E.E., Kanouse D.E., Kelley H.H., Nisbett R.E., Valins, S.
- Kohlberg L. (1984). Essays on moral development, *The psychology of moral development*. Vol. 2. San Francisco: Harper and Row.
- Krampe G. (1986). *Handlungsleitende Kognitionen von Lehrern. Instrumentalitätstheoretische Vorhersagen pädagogischer Präferenzen* [Teachers' action-guiding cognitions. Instrumental-theoretical predictions of educational preferences]. Göttingen: Hogrefe.
- Lengenfelder P. (2003). *Konstruktion und Evaluation eines Fragebogens zur Beurteilung universitärer Lehrveranstaltungen durch Studierende* [Construction and evaluation of a questionnaire for the student ratings of university courses]. Unpublished Masters Thesis, University of Salzburg.
- Lengenfelder P., Patry J.-L., Schimke H. (2001). Bericht zur Beurteilung der Lehrveranstaltungen durch die Studierenden an der Universität Salzburg [Report on the student ratings of university courses at the University of Salzburg]. University of Salzburg. URL: http://www.sbg.ac.at/evaluation/eval_bericht01.doc (June 30, 2003).
- Lengenfelder P., Patry J.-L., Baumann U., Paschon A., Ehmer J., Schimke H. in press: *Globale und additive Erhebungen studentischer Beurteilungen universitärer Lehrveranstaltungen – eine Multitrait-Multimethod Analyse* [Globale and additive assessments of student ratings of university teaching]. To be published in *Empirische Pädagogik*.
- Marsh H.W. (1991). Multidimensional students' evaluations of teaching effectiveness: A test on alternative higher-order structures. *Journal of Psychology*, 83, p. 285-296.
- Marsh H.W., Ware J.E., jr. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *Journal of Educational Psychology*, 74, p. 126-134.
- Nitz H.-R. (1993). Am runden Tisch und zwischen den Stühlen. Zur Position des Helfers bei der Interaktion psychosozialer Systeme [At the round table and between the chairs. About the position of the helping person in interactions of psycho-social systems]. In: Kaiser, P. (Ed.). *Psycho-Logik helfender Institutionen. Beiträge zu einer besseren Nutzerfreundlichkeit der Organisationen im Sozial- und Gesundheitswesen*. Heidelberg: Asanger, p.145-170.
- Paschon A., Thonhauser J. (2002). Rezension von "Deutsche Gesellschaft für Evaluation (Wolfgang Beywl): Standards für Evaluation. [Review of „German society for evaluation“ [W. Beywl]: Standards for evaluation] Köln: Eigenverlag". *Pädagogisches Handeln*, 6, p. 328-330.
- Patry J.-L. (1991). *Transsituationale Konsistenz des Verhaltens und Handelns in der Erziehung* [Transsituational consistency of action and behavior in education]. Bern: Lang.
- Patry J.-L. (1997a): Eine Person - mehrere Werte. Überlegungen zum intrapersonalen Wertpluralismus [One person – several values. Reflections on the intrapersonal pluralism of values]. *Pädagogische Rundschau*, 51, p. 63-81.
- Patry J.-L. (1997b): The lesson interruption method in assessing situation-specific behavior in classrooms. *Psychological Reports*, 81, p. 272-274.
- Patry J.-L. (2002). Bewertung von Lehrveranstaltungen durch Studierende - Möglichkeiten und Grenzen der Aussagen [Student ratings of teaching – possibilities and limits of the statements]. Paper read at the symposium „Evaluation universitärer Lehre: Quo Vadis?“, Vienna University, February 28, 2002.
- Rindermann H. (2001). Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierten Unterrichts [Evaluation of teaching. Introduction and survey concerning science and practice of the evaluation of lectures at universities]. Landau: Empirische Pädagogik.

Shulman J.H. (1990): Now you see them, now you don't: Anonymity versus visibility in case studies of teachers. *Educational Researcher*, 19 (6), p. 11–15.

Slavin R.E. (1986). *Educational Psychology: theory into practice*. New Jersey: Prentice-Hall.

Smith M.L., Glass G.V. (1987). *Research and evaluation in education and the social sciences*. Boston: Allyn and Bacon.

Smith S.W., Medendorp C. L., Ranck S., Morrison K., Kopfman J. (1994). The prototypical features of the ideal professor from the female and male undergraduate perspective: The role of verbal and nonverbal communication. *Journal on Excellence in College Teaching*, 5(2), 5–22. URL <http://www.lib.muohio.edu/ject/ject5-2/smith5-2.html> (September 25, 2003).

Task Force "Evaluation von Lehrveranstaltungen an der Universität Salzburg" 2002: Lehrveranstaltungs-Rückmeldung

(Course feedback). URL <http://www.sbg.ac.at/evaluation/> (October 29, 2003).

Tatro C.N. (1995): Gender effects on student evaluations of faculty. *Journal of Research and Development in Education*, 28, p. 169–173.

Thiel C. (1980). Dilemma (Dilemma). In: J. Mittelstraß (Hrsg.), *Enzyklopädie Philosophie und Wissenschaftstheorie*, Band 1: A–G, Mannheim: Bibliographisches Institut, p. 482–483.

UOG (1993). Bundesgesetz über die Organisation der Universitäten [Law about the organisation of the universities]. *Bundesgesetzblatt* 805.

Watzlawick P., Beavin J., Jackson D.D. (1969). *Human communication: Forms, disturbances, paradoxes*. Oxford, England: Hans Huber Publishers.

*Įteikta 2004 m. kovo mėn.
Delivered 2004 March*

Prof. dr. JEAN-LUC PATRY
Zalcburgo universiteto
Edukologijos katedros profesorius.
Tyrimų interesų sritys:
situacijų specifiškumas, ugdymo vertybės, įvertinimas,
teorijos ir praktikos ryšys, konstruktyvizmas ir kt.

*Zalcburgo universitetas
Edukologijos katedra
Akademiestrasse 26, A-5020 Salzburg, Austria
jean-luc.patry@sbg.ac.at*

Prof. Dr. JEAN-LUC PATRY
Professor at the Department of Education
at the University of Salzburg.
Research interests:
situation specificity, values education, evaluation,
theory-practice relationship, constructivism, etc.

*University of Salzburg
Department of Education
Akademiestrasse 26, A-5020 Salzburg, Austria
jean-luc.patry@sbg.ac.at*

Dr. ANGELA GASTAGER
Zalcburgo universiteto
Edukologijos katedros asistentė
Tyrimų interesų sritys:
subjektyvistinės teorijos, įvertinimo dilemų struktūra, suaugusiųjų švietimas: nuosaikūs dėstymo ir studijavimo aplinkos konstruktyvizmas, komandinio darbo grupėje kokybė.

*Zalcburgo universitetas
Edukologijos katedra
Akademiestrasse 26, 5020 Salzburg, Austria
angela.gastager@sbg.ac.at*

Dr. ANGELA GASTAGER
Assistant at the Department of Education
at the University of Salzburg.
Research interests:
subjective theories, dilemma structure of evaluation,
adult education: (moderate) constructivist teaching-and-learning-settings, quality of learners' teamwork.

*University of Salzburg
Department of Education
Akademiestrasse 26, 5020 Salzburg, Austria
angela.gastager@sbg.ac.at*