

Jean-Luc PATRY

Zalcburgo universitetas • University of Salzburg

KRITINIO MULTIPLIZMO¹ PROBLEMOS ĮVERTINIMO TYRIMUOSE: TEORIJŲ MULTIPLIZMAS IR ŠALIŠKUMO ANALIZĖ

ISSUES IN CRITICAL MULTIPLISM IN EVALUATION RESEARCH: MULTIPLISM OF THEORIES AND ANALYSIS OF BIASES

SANTRAUKA

Kokybės užtikrinimo ir įvertinimo tyrimai, kaip ir kiti socialiniai tyrimai bei jų taikymas, kelia daugybę problemų. Šiame straipsnyje pirmiausiai norėčiau pateikti tokių problemų sąrašą, kuris vis dėlto liks nepilnas. Straipsnyje teigiama, kad nors ir nėra „tobulo“ šių problemų sprendimo būdo, kritinis multiplizmas galėtų būti toks požiūrių modelis, kuris leistų sumažinti tas problemas arba bent padėtų jas išsiaiškinti ir mėginti spręsti. Kritinis multiplizmas - tai bendro pobūdžio metodologinis modelis, kurį nuodugnai išsiaiškinti čia būtų neįmanoma; norėčiau aptarti tik kai kuriuos svarbiausius klausimus ir remdamasis jais pateikti keletą pasiūlymų, kaip turėtų būti atliekami socialiniai tyrimai.

PAGRINDINIŲ SĄVOKŲ APIBRĖŽIMAI

- *Diskriminantinis pagrįstumas* – kai du skirtingų savybių vertinimai tais pačiais ar skirtingais metodais duoda skirtingus rezultatus ir tuo parodo, jog tos dvi savybės iš tikrųjų yra skirtingos.
- *Gabumas* – „Bet kuri asmens savybė, kuri prognozuoja sėkmės tikimybę taikant tam tikrą poveikį. (...) Asmenybė kaip ir gebėjimas įtakojantis reagavimą į tam tikrą mokymą.“ (Cronbach, Snow, 1981, p.5).
- *Įvertinimas* – praktinio konteksto analizė mokslinėmis priemonėmis, kuriomis siekiama surasti sprendimus tame kontekste; (1) *institucijų įvertinimas* – tyrimas, ar institucija pasiekia išskeltus tikslus ir kaip jos darbas galėtų būti patobulintas; (2) *programų įvertinimas* – tyrimas, ar tam tikra programa tinkama pasiekti išskeltus tikslus ir kokiomis sąlygomis ji tinkama naudoti.
- *Įvertinimo tyrimas* (1) – įvertinimo arba įvertinimo studijų tyrimas; (2) įvertinimo atlikimas.
- *Konvergentinis pagrįstumas* – vienodi dviejų tos pačios savybės vertinimų rezultatai, taikant skirtingus metodus.
- *Kritinis multiplizmas* – pripažinimas, kad socialiniuose moksluose kiekvienas požiūris dažniausiai turi savo teigiamų ir neigiamų pusių, todėl tam tikro vieno iš jų pasirinkimas sąlygoja teorinį ar metodinį šališkumą. Kritinis multiplizmas – tai metateorinis modelis, rodantis, kad naudojami sistemingai suplanuoti daugialypiai požiūriai, siekiant nustatyti šališkumą arba netgi jo išvengti.

ABSTRACT

Quality assurance and evaluation research, like other fields of social research and its application, are confronted with a series of problems. In the present paper, I want first to give a list of such problems, although necessarily incomplete. It is then claimed that while there is no “perfect” solution to these problems, critical multiplism may be a set of approaches which might attenuate the problems or at least make them more visible so that one can deal with them. Critical multiplism is a general methodological framework which cannot be fully accounted for here; rather, I want to present some major issues and on their base make some suggestions about how to conduct social research.

DEFINITIONS OF KEY WORDS

- *Discriminant validity* – two assessments of different traits with the same or with different methods show different results, thus showing that indeed the two traits are distinct.
- *Aptitude* – “(A)ny characteristic of a person that forecasts his probability of success under a given treatment. (...) Personality as well as ability influences response to a given kind of instruction” (Cronbach & Snow, 1981, p. 5).
- *Evaluation* – analysis of a practical context with scientific means in order to make decisions in this practical context; (1) evaluation of institutions: analysis whether the institution achieves its goals and how it could be improved; (2) evaluation of programs: analysis whether a given program is appropriate to achieve its goals and under what conditions it is appropriate to use it.
- *Evaluation research* – (1) research on evaluation or on evaluation studies; (2) doing evaluation.
- *Convergent validity* – two assessments of the same trait with different methods show the same result.
- *Critical multiplism*: Since each approach (in social science) tends to have both strengths and weaknesses, the choice of a particular one introduces theoretical or methodological biases. Critical multiplism is a meta-theoretical framework suggesting that multiple approaches are used, yet systematically planned to identify and possibly even rule out such biases.

¹ „Kritinio multiplizmo“ sąvoka kol kas neturi savo reglamentuoto lietuviško atitikmens. „Kritinis multiplizmas“ galėtų būti verčiamas kaip „kritinis daugialypumas“, „kritinis įvairumas“, „kritinio įvairumo metodika“ ar pan. Šiame straipsnyje jis verčiamas tiesiogiai.

- *Pagrįstumas arba tinkamumas* – „išvados teisingumas arba tikslumas, arba patvirtinimo laipsnis“ (Shadish ir kt., 2002, p. 513); didelis pagrįstumas rodo nežymią sistemine klaidą; pagrįstumui patikrinti galima naudoti Campbell ir Fiske (1959, p. 83) sąvoką: „Pagrįstumas žymi atitikimą tarp dviejų bandymų pamatuoti tą pačią savybę visiškai skirtingais metodais“.

- *Patikimumas arba tikslumas* – tam tikros charakteristikos vertinimo tikslumas, kad ir kokie būtų to vertinimo rezultatai; didelis patikimumas rodo nežymią atsitiktinę paklaidą; norint patikrinti patikimumo lygį, galima pasiremti Campbell ir Fiske (1959, p. 83) sąvoka: „Patikimumas – tai atitikimas tarp dviejų bandymų pamatuoti tą pačią savybę visiškai panašiais metodais“.

- *Pigmalionas* – iš graikų mitologijos žinomas Pigmalionas buvo Kipro karalius, kuris pamilo savo paties sukurtą statulą; Artemidė įkvėpė gyvybę statulai ir Pigmalionas ją vedė. Rosenthal ir Jacobson (1968), atradę šį efektą (dėstytojai studentų sėkmei įkvėpia gyvybę), pavadino jį *Pigmaliono efektu* (arba *Pigmalionu auditorijoje*).

- *Šališkumas* – sistemingos klaidos, keliančios pavojų (įvertinimo) pagrįstumui.

ĮVADAS

Mažiausiai 80 metų kalbama, kad socialiniai mokslai išgyvena krizę (pvz., Vygotsky, 1927; Mertens, Fuchs, 1978). Kai atliekant tyrimus ir įvertinimus į socialinį elgesį sutelkiamas ypač didelis dėmesys, susiduriama su daugybe problemų. Kai kurios jų išvardijamos žemiau:

- Priemonės, skirtos socialinio elgesio vertinimui, ypač klausimynai, yra nelabai patikimi ir nevisai pagrįsti (plg. Mischel, 1968). O tai rodo, kad vertinimai nėra tikslūs ir dažnai nežinoma, kas iš tiesų yra matuojama. Toks dalykas kelia daug pavojų bet kuriam tokiu tyrimu pagrįstam sprendimui (o tai ir yra įvertinimo tikslas), nes pats sprendimo priėmimo pagrindas kelia abejonių.

- Atskirų tyrimų rezultatai paprastai neatsikartoja, jeigu pakartotinis tyrimas šiek tiek skiriasi nuo pirminio originalaus tyrimo, netgi ir tuo atveju kai skirtumai nežymūs; pvz., įvertinus šešių atsitiktinai parinktų pakartotinių tyrimų rezultatus, beveik 55 proc. iš 64 rezultatų parodė „duomenų pasikartojimo stoką“ (Amir, Sharon, 1991, p. 63); tačiau pakartotinių tyrimų kol kas stokojama (ten pat; kt., pvz., Karr, Larson, 2005), juos sudėtinga paskelbti (Bornstein, 1991; Neuliep, Crandall, 1991), ir tai ypač pasakytina apie nesėkmingus pakartotinius tyrimus, kuriems taikomas „numarinimo stalčiuje“ principas (Rosenthal, 1979).

- Tyrinėtojai linkę pervertinti savo teorijų pagrįstumo ribas (Patry, 1991; Amir, Sharon, 1991); tai tiesa apskritai paėmus, bet ypač svarbu kai imama kalbėti apie įvertinimus, nes jų pagrįstumo ribos nėra plačios, o tas teorinis pagrindas, kuris šiuo atveju padėtų, dažniausiai per menkas; visa tai taip pat taikytina ir ateities prognozėms, kurioms įvertinimo procesuose skiriamas pagrindinis dėmesys.

- Jeigu ieškant atsakymo į tą patį klausimą buvo atlikti skirtingi tyrimai, tai kuo tikslesni ir labiau tinkami metodai taikomi (pvz., kiekybiniai metodai tikslesni už kokybinius), tuo mažesnė darosi tikimybė gauti reikšmingų rezultatų. Tai, pvz., akivaizdu, lyginant mokslinio simpoziumo metu pateiktų skirtingų tyrimų rezultatus; vieną kartą esu daręs tai neformaliu būdu, bet nesu girdėjęs apie sistemingus tyrinėjimus.

- Ryšys tarp teorijos ir praktikos, kuris laikomas vienu svarbiausių klausimų socialiniuose moksluose apskritai, o

- *Validity* – “the truth of, correctness of, or degree of support for an inference” (Shadish et al., 2002, p. 513); high validity means little systematic error; to check for validity, one can use the concept by Campbell and Fiske (1959, p. 83): “Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods.”

- *Reliability* – how precise is the assessment of a particular characteristic, whatever it is; high reliability means little random error; to check for reliability, one can use the concept by Campbell and Fiske (1959, p. 83): “Reliability is the agreement between two efforts to measure the same trait through maximally similar methods”.

- *Pygmalion* – in Greek mythology, Pygmalion was a king of Cyprus who fell in love with a statue of his own making; Artemis gave life to the statue, and he married her. Rosenthal and Jacobson (1968) called the effect they found (teachers “make” students success come to life) the “Pygmalion effect” (or “Pygmalion in the classroom”).

- *Bias* – Systematic error jeopardizing validity.

INTRODUCTION

At least for 80 years, social sciences have been claimed to be in a crisis (e.g., Vygotsky, 1927; Mertens & Fuchs, 1978). Particularly when social behavior is at stake, research and evaluation encounter many problems. Some of them are the following:

- The instruments to assess social behavior, particularly questionnaires, have low reliability and validity (cf. Mischel, 1968), which means that the assessments cannot be precise and that one does not really know what is measured. This jeopardizes particularly any research-based decision (which is the aim of evaluation) because its base is doubtful.

- Results from particular studies are usually not replicable if the replication study is slightly different from the original one, even if the changes are judged to be irrelevant; e.g., in the replications of six randomly chosen studies, almost 55% of the 64 results indicate “a lack of any replicability of data” (Amir & Sharon, 1991, p. 63); however, replication studies are underutilized (op. cit.; another example is Karr & Larson, 2005), they are difficult to publish (Bornstein, 1991; Neuliep & Crandall, 1991), and this is even more true for failed replications according to the “file drawer effect” (Rosenthal, 1979).

- Researchers tend to overestimate the domain of validity of their theories (Patry, 1991; Amir & Sharon, 1991); this applies in general but particularly in evaluations because these are not designed for broad domains of validity and the theoretical background which would help in this regard is mostly insufficient; this holds also for predictions for the future which are the main focus of evaluations.

- If different studies are done to answer the same question, the more strict and appropriate the methods are (e.g., quantitative methods are more strict than qualitative ones), the less likely it is to find significant results. This can be seen, e.g., when comparing the results of different studies presented within a scientific symposium; I did this once informally, but I am not aware of a systematic study.

ypač įvertinimuose, toli gražu nėra taip išplėtotas, kad praktiškai galima būtų vertinti kaip teorijos pritaikymą (plg., pvz., Bunge, 1967 arba iš kitos perspektyvos Gage, 1978), išskyrus Lewin (1951, p. 196) tvirtinimą, kad nėra nieko praktiškiau už gerą teoriją.

- Socialiniuose tyrimuose daug artefaktų ir šališkumo (Rosenthal, Rosnow, 1969; Bungard, Bay, 1982; ir daugelis kitų), t.y. savo vertinimuose ir išvadose mes darome sistemingas klaidas.

- Įvertinimų procesuose dalyvauja daugybė socialinių partnerių, keliančių skirtingus, dažnai prieštarigus, reikalavimus (Patry, Gastager, 2004).

- Tyrinėtojai pakliūva į teorijų patvirtinimo ir jų paneigimo spąstus. Viena vertus, remiantis kritinio racionalizmo koncepcija (Popper, 1934/1976; plg. žemiau) tikslas yra paneigti teorijas. Kita vertus, „atskiriems mokslininkams, kaip ir visiems žmonėms, geriau pavyksta pastebėti savo teorijų patvirtinimą, o ne paneigimą (Faust, 1984; Mahoney, 1976)“ (Shadish, 1986, p. 79). Paprastai tyrimo tikslas nėra paneigti kažkieno teoriją, kuria remtasi; dažniau nusivilinama tada, kai nepavyksta įrodyti, jog mūsų hipotezės pasitvirtino.

Paminėtina ir mokslinių tyrimų apskritai, o ypač socialinių mokslų tyrimų kritika metateoriniame lygmenyje, t.y., mokslo filosofijos arba filosofinės epistemologijos lygmenyje. Čia išvardijama keletas kritikuojamų dalykų:

- Savo mokslinių revoliucijų koncepcijoje Kuhn (1962) pabrėžia, kad mokslas nesiplėtoja tolygiai, bet patiria lūžius ir revoliucijas, tačiau socialiniai mokslai nėra pasiekę net šio etapo, greičiau tik ikimokslinį.

- Feyerabend (1975) nuomone, idėja, kad mokslas gali ir turėtų būti plėtojamas pagal nekintamas ir universalias taisykles, per daug naivi ir žalinga mokslui. Visos metodikos turi savo trūkumų. Jo nuomone, „mokslas yra iš esmės anarchistinis sumanymas: teorinis anarchizmas yra humaniškesnis ir tinkamesnis pažangai skatinti negu jo įstatymų ir tvarkos alternatyvos“ (ten pat, p. 17), tad jis kelia klausimą, ar turime „iš tikrųjų patikėti, kad naivios ir kvailokos taisyklės, kuriomis vadovaujasi mokslininkai, gali išaiškinti tokį „sąveikų labirintą“ (p. 17 ir t.t.), kur pastaroji sąvoka apima „žmogaus kaitos sudėtingumą ir nenuspėjimą bet kokio veikėjo ar žmogaus sprendimo galutinių padarinių pobūdį“ (ten pat, p. 17).

- Glasersfeld (1995) teigia, kad nėra jokio būdo sužinoti tiesą arba „kas yra iš tikrųjų“. Greičiausiai, viskas, ką žinome, yra mūsų interpretacija (mūsų išgalvojimas, sako Foerster, 1998). Ne tiesa, o pasiekiamumas turėtų būti dėmesio centre, be to, yra įvairių pasiekiamumo kriterijų (Patry, 2001a).

- Remiantis sistemų teorija (pvz., Watzlawik ir kt., 1974) galima manyti, kad viskas susiję su viskuo; negalima izoliuoti kokios nors vienos sistemos ir įsivaizduoti, kad nustatėme, kaip ji iš tikrųjų veikia. Kiti autoriai yra teigę, kad tikrovė pernelyg sudėtinga, kad ji būtų suprasta, naudojantis mokslinėmis (kiekybinėmis) priemonėmis (pvz., Cziko, 1989).

- Popper (1934/1976) pabrėžė, kad indukcinis metodas yra neįmanomas. Tačiau jo pasiūlytas sprendimas – paneigimas, o ne patvirtinimas – yra neįmanomas, kai teiginiai yra tikimybiniai, o tokie teiginiai dažni socialiniuose moksluose.

- The relationship between theory and practice, one of the most important issues of social science in general and of evaluation in particular, is far from being developed in such a way that practice can be seen as application of theory (cf., e.g., Bunge, 1967, or, from another perspective, Gage, 1978), despite Lewin's (1951, p. 196) claim that there is nothing as practical as a good theory.

- There are many artefacts or biases in social research (Rosenthal & Rosnow, 1969; Bungard & Bay, 1982; and many more), i.e., we make systematic errors in our assessments and inferences.

- In evaluations, there are many stakeholders with many different requirements, some of which are contradictory (Patry & Gastager, 2004).

- Researchers are caught in a confirmation-refutation trap. On one hand, according to the concept of critical rationalism (Popper, 1934/1976; cf. below), their aim is to refute theories. On the other hand, “individual scientists, just like all human beings, are better at noting confirmations of their theories than disconfirmations (Faust, 1984; Mahoney, 1976)” (Shadish, 1986, p. 79). Typically, the aim of doing a study is not to refute one's theory which usually is something one believes in; rather, we are disappointed if we do not succeed to show that our hypotheses are confirmed.

In addition, there have been several criticisms of scientific research in general and of social science research in particular on a meta-theoretical level, i.e., from the standpoint of philosophy of science or philosophical epistemology. Some of these criticisms are:

- With his concept of scientific revolutions Kuhn (1962) emphasizes that science is not growing steadily, but that there are ruptures and revolutions – however, social sciences are not even in that stage, but rather “pre-scientific”.

- Feyerabend (1975) argues that the idea that science can, and should, be run according to fixed and universal rules is both too simple and detrimental to science. All methodologies have their limitations. For him, “science is an essentially anarchistic enterprise: theoretical anarchism is more humanitarian and more likely to encourage progress than its law-and-order alternatives” (op. cit., p. 17), and he asks whether we are “really to believe that the naïve and simple-minded rules which methodologists take as their guide are capable of accounting for such a ‘maze of interactions’” (op. cit., pp. 17f.), the latter term referring to the “complexity of human change and the unpredictable character of the ultimate consequences of any given actor or decision of men” (op. cit., p. 17).

- Glasersfeld (1995) criticizes that there is no way we can know truth or “what is the case”. Rather, all we can know is our own construction (our invention, says Foerster, 1998). Instead of truth, viability should be at stake, and there are different viability criterions (Patry, 2001a).

- Based on systems theory (e.g., Watzlawik et al., 1974) one can assume that everything is connected with everything; one cannot isolate one system and pretend to have identified how it really works. Others have said that reality is too complex to catch it with scientific (quantitative) means (e.g., Cziko, 1989).

- Popper (1934/1976) has pointed to the impossibility of induction. His proposed solution – refutation instead of verification –, however, is impossible if the statements are probabilistic, which is usually the case in social sciences.

- Theoretical statements are abstract. The more concrete they are, the smaller are the respective domains of

- Teoriniai teiginiai yra abstraktūs. Kuo jie konkretesni, tuo siauresnės atitinkamos pagrįstumo ribos; kuo platesnės teiginio pagrįstumo ribos, tuo jis abstraktesnis (Herrmann, 1979, p. 160–161). Tai abstraktumo ir konkretumo dilema.

- Teorijos paprastai negali gerai paaiškinti socialinio elgesio situacijos specifiškumo, o tai kelia daugybę problemų (Patry, 2001b).

Galima būtų išvardinti ir dar daugiau kritinių pastabų dėl socialinių mokslų ir jų tyrimų. Atrodo, tarsi tyrimus iš visų pusių suptų pavojai, tad kyla klausimas, ar socialiniai mokslai kada nors galės duoti naudingų rezultatų. Iš tikrųjų palyginus su kitais mokslais, pvz., fizika, chemija ar netgi biologija, matyti, kad socialiniai mokslai ne tokie tikslūs (plg., pvz., Hedges, 1987) ir mažiau apibrėžti.

Šios problemos daro įtaką ir įvertinimo tyrimams bei kokybės užtikrinimui. Visos aukščiau išsakytos kritinės pastabos gali tikti ir, pavyzdžiui, programų vertinimams siekiant patobulinti dėstyto kokybę. Geras bandymas, kuriuo čia pasiremti negalime, būtų pasirinkti bet kurį tyrimą ir jį išanalizuoti remiantis aukščiau pateiktomis kritinėmis pastabomis. Daugeliu atvejų tik vienas kitas rezultatas liktų galioti.

Socialiniuose tyrimuose egzistuoja keturi būdai čia aptartoms problemoms spręsti; (a), (b) ir (d) paimta iš Bungard ir Bay (1982), (c) sąlygoja bendroji patirtis:

(a) Problemas galima ignoruoti ir manyti, kad socialiniai mokslai neišgyvena jokios krizės. Tai dažniausia reakcija, kuri psichologiškai suprantama, tačiau problemų neigimas niekada nebuvo teisingas sprendimas, jis negali padėti nei mokslui žengti į priekį, nei išspręsti egzistuojančių problemų.

(b) Problemos galima pamėginti išvengti renkantis tariamai nepriekaištingą metodologinę nuostatą. Pavyzdžiui, siekdami socialinio pobūdžio tyrime išvengti artefakto problemos, kylančios dėl subjekto reaktyvumo (žinodami, kad yra stebimi, subjektai elgiasi kitaip), tyrinėtojai pasitelkia reaktyvumo nesukeliančius metodus (Webb ir kt., 1981). Toks pasirinkimas galėtų padėti išspręsti vieną problemą, tačiau kitos neišnyksta, o joms gali būti taikomi tie patys argumentai kaip ir (a) atveju.

(c) Mokslinių metodų priimtumą galima apskritai paneigti ir rinktis kitus požiūrius, pvz., ezoterizmą. Tačiau tai nėra geriausias būdas, nes susidūrimas su problemomis nėra priežastis atsisakyti tyrimų ir griebtis kitos tariamai malonesnės veiklos. Moksliniai tyrimai gali turėti problemų; tačiau kad ir kokie būtų jų trūkumai, iš viso to ką turime jie neabejotinai lieka geriausi ir tinkamiausi pažinimo būdai palyginti su visais kitais būdais, kuriais siūloma tyrimus pakeisti.

(d) Problemas galima mėginti išaiškinti teoriniame ir metodiniame lygmenyje, ir jeigu jos neišsprendžiamos, tai bent jau suvokiamos ir pritaikomi principai, kurie padeda (i) nustatyti, ar problema egzistuoja ir (ii) apriboti jos poveikį. Kaip matysime vėliau, kritinis multiplizmas šiuo atveju gali pasiūlyti potencialius būdus tyrimus atlikti geriau negu jie buvo atliekami iki šiol.

Iš aukščiau pateiktų pasiūlymų (d), be abejonės, yra priimtinausias. Galima suformuluoti tokią pagrindinę taisyklę:

1 pasiūlymas: Pripažinti socialinių mokslų tyrimų problemas, bet interpretuoti jas kaip iššūkius ir kiek įmanoma jas sumažinti!

validity; the broader the domain of validity of a statement is, the more abstract it is (Herrmann, 1979, pp. 160-161): This is the generalizability-concreteness dilemma.

- Theories usually do not account sufficiently for situation specificity of social behavior, which creates many problems (Patry, 2001b).

One could mention many more criticisms to social science and research. It seems that research is jeopardized from many sides, and one wonders whether social science will ever be able to bring about useful results. Indeed, when compared with other scientific disciplines such as physics or chemistry or even biology, one must say that the social sciences are much softer (cf., e.g., Hedges, 1987) and fuzzy.

These problems have their impact also on evaluation research and quality assurance. All of the above criticisms can be applied, for instance, to assessments within programs to improve quality of teaching. It would be a good exercise – which cannot be done here – to take any piece of research and to check whether it holds against the criticisms presented above. In most cases, very little of the results will survive.

In social research, there have been four ways to deal with these kinds of problems; (a), (b), and (d) are from Bungard and Bay (1982), (c) is a common experience:

(a) One can just ignore the problems and presume that there is no crisis in social research at all. This is the most frequent reaction. This reaction is psychologically understandable, yet denial of problems is never a good reaction, and it will certainly contribute neither to the progress of the science nor to solving the problems.

(b) One can try to avoid problems by using presumably flawless approaches. For instance, to avoid the artifact problem in social research due to the subjects' reactivity (the subjects' knowing that they are being observed makes them behave differently), researchers use nonreactive methods (Webb et al., 1981). Such a reaction might help to solve one problem, the others, however, are not dealt with, and for them, the same argument holds as for (a).

(c) One can deny the appropriateness of scientific methods at all and to turn to other approaches such as esotericism. This is not appropriate since the existence of problems is not a reason to give up research and to turn to presumably more gratifying activities. Research may have its problems; however, as bad as it may be yet, it is still the best that is available and definitely more appropriate than all approaches which have been proposed as substitutes.

(d) One can try to account for the problems on theoretical and methodological levels and if not solve the problems at least be aware of them and apply principles that permit (i) to identify whether there are problems and (ii) to minimize their impact. As will be discussed below, critical multiplism can be an approach in this sense with the potential to do better than research until now.

From the suggestions above, (d) is certainly the most appropriate. As a general rule, we can formulate the following

suggestion 1: Acknowledge the problems of social research, but interpret them as challenges and do your best to reduce them to a minimum!

1 KRITINIS MULTIPLIZMAS

1.1. KRITINIO MULTIPLIZMO PRINCIPAI IR LOGINIS PAGRINDAS

Kiekvienas požiūris – vertinimo priemonė, metodas, teorija, kt. – yra šališkas, turi savo „silpnų vietų“ arba „tamsių dėmių“, iš kurių keletas paminėta įvade pateiktame problemų sąrašė. Kai vadovaujamosi tik vienu požiūriu, šališkumas užgožiamas ir jį sunku pastebėti. Būtent čia kritinio multiplizmo koncepcijos taikymas gali padėti.

Trumpai tariant, kritinis multiplizmas reiškia, kad naudojami daugialypiai požiūriai atsakyti į tam tikrą (tyrimo) klausimą. Tačiau visai nereikalingas neprotingas aktyvumas, kai „bet kas tinka, svarbu padaryti kiek įmanoma daugiau“. Kritinis multiplizmas yra kritinis sisteminis būdas, leidžiantis apimti kuo daugiau įvade iškeltų problemų, kad jų išvengti (Shadish, 1994). Hetherington (1997) pasiūlė, kaip atskirti kritinį multiplizmą nuo kitų būdų (1 lentelė); jis pasirinko „apmaštyta – neapmaštyta“ skirtumu analizuodamas metodus, bet tai gali būti taikoma ir kitoms sritims, pavyzdžiui, teorijoms.

Kritinis multiplizmas laikosi nuostatos, kad nėra nė vieno požiūrio, kuris būtų laikomas geresniu už kitus;

1 CRITICAL MULTIPLISM

1.1. PRINCIPLES OF CRITICAL MULTIPLISM AND ITS RATIONAL

Every single approach – assessment tool, method, theory, etc. – has its biases, its “black spots”, some of which are pointed at in the list of problems mentioned in the first section. Using only a single approach will obscure these biases so that they cannot be recognized. This is where critical multiplism can provide a remedy.

Briefly stated, critical multiplism means that to use multiple approaches to answer a particular (research) question, yet not in the sense of a mindless activism of the type “anything goes, just do as much as possible”, but rather in a critical and systematic way so that as many problems among those mentioned in section 1 and others are accounted for and – possibly – ruled out (Shadish, 1994). Hetherington (1997) proposed the systematic as in table 1 to distinguish critical multiplism from other approaches; he applied this distinction to methods, but it can be extended to other domains, like theories.

Critical multiplism means that there is not one single best approach; rather, each method has its advantages and

1 lentelė. **Kritinio multiplizmo ir kitų požiūrių skirtumai** (pagal Hetherington, 1997, 18 skaidrė)
Table 1: **Distinction of critical multiplism from other approaches** (from Hetherington, 1997, slide 18)

	NEAPMAŠTYTA THOUGHTLESS	APMAŠTYTA THOUGHTFUL
PAVIENIS METODAS SINGLE METHOD	Beprasmis monizmas Skurdi teorija Žemas pagrįstumas	Tikslus monizmas Niutono mokslas MAX-MIN-CON ²
DAUGIALYPIAI METODAI MULTIPLE METHODS	Beprasmis daugialypumas Skurdi teorija Bet kas tinkama	Kritinis multiplizmas
	Mindless Multiplism Poor science Anything Goes	Critical Multiplism

² MAXMINCON principą suformulavo Kerlingeris. Tai yra klasikinis eksperimentinio tyrimo principas. Jis reiškia didink kiek įmanoma sistemų skirtingumą. Mažink kiek įmanoma paklaidų skirtumą. Kontroluok klaidinančius kintamuosius. Idėja yra ta, kad siekiama didelių skirtumų tarp eksperimentinės ir kontrolinės grupių kaip parenkamų sistemų (t.y., procedūros turi būti veiksmingos), maži skirtumai grupių viduje (homogeninės grupės, aukšto patikimumo ir tikslumo vertinimo instrumentai) ir kad kintamieji, kurie gali sukelti ženklias klaidas turi būti kontroliuojami (t.y. bet koks veiksnys, kuris gali daryti papildomą įtaką, turi būti eliminuotas). Tai būdinga laboratoriniams eksperimentams, kuriais galima aiškiai parodyti, kad vienas specifinis nepriklausomas kintamasis (kokia nors veikla) turi įtakos kokiam nors priklausomam kintamajam (kuris yra matuojamas). Realiam gyvenime vis dėl to negalima eliminuoti visas įtakas, taigi maxmincon teigia, kad eksperimentai stokoja (ekologinio) pagrįstumo, t.y. tokių eksperimentų išvados ar teiginiai neatitinka to, kas vyksta realiam gyvenime.

² MAXMINCON principle stems from Kerlinger and is classical principal from experimental research; it means Maximize systematic variance. Minimize error variance. Control for confounding variables. The idea is that one should look to get high differences between experimental and control groups (i.e., powerful treatments), little variance within the groups (homogeneous groups, highly reliable assessment instruments), and that variables which might produce systematic errors should be controlled (i.e., any factor that might have additional influences should be eliminated). This yields laboratory experiments in which one can powerfully demonstrate that one specific independent variable (treatment) has an influence on one specific dependent variable (which is measured). However, life is not such that one can eliminate other influences etc., so the maxmincon – experiments lack ecological validity, i.e. the statements from such experiments do not permit statements about what is happening in real life.

manoma, jog kiekvienas metodas turi savo pranašumų ir trūkumų. Todėl siūloma taikyti keletą metodų vienu metu. Jeigu taikant skirtingus metodus gaunami panašūs rezultatai, galima tikėti jų pagrįstumu, kol nėra priešingų įrodymų. Jeigu rezultatai labai skiriasi, akivaizdu, kad yra problemų: žinome, kad rezultatais pasitikėti negalime, bet nežinome, kurie iš jų patikimesni lyginant su kitais, ir kuriuo atžvilgiu jie laikytini patikimesniais.

Žemiau pateikiami kritinio multiplizmo pavyzdžiai (taip pat žr. Houts ir kt., 1986; Patry, 1989; Shadish, 1986 ir kt.):

- Daugialypis operacionalizmas ir daugialypiai metodai (Campbell, Fiske, 1959);
- Sudėtinių daugiamacių modelių analizės taikymas vietoj vienamacių modelių (tai susiję su konstruktais, kurie bus tiriami);
- Daugialypių konkuruojančių hipotezių patikrinimas (Campbell, 1969);
- Daugialypių susijusių tyrimų taikymas, pavyzdžiui, tyrimų programose (Herrmann, 1976), tyrimų strategijose, apimančiose keletą skirtingai planuotų vertinimų tokių, kaip pvz., įvairiapusiai tyrimai (Seashore, Louis, 1984; Herriott, Firestone, 1983; Slavin, Madden, 1993; Audet, d'Amboise, 2001) ir kt.
- Skirtingų nepriklausomai atliktų tyrimų rezultatų sintezė (pvz., metaanalizių būdu; Cook, Leviton, 1980; Cook, 1991; Glass, McGaw, Smith, 1981; Rosenthal, 1989; ir t.t.);
- Tų pačių duomenų nagrinėjimas taikant skirtingus statistinius metodus;
- Skirtingų teorijų taikymas hipotetiniams atsakymams į panašius tyrimų klausimus gauti ir rezultatams interpretuoti (teorijų konkurencija, koegzistencija ir netgi jų papildomumas, žr. pvz., Patry, 1998).

Kaip minėta aukščiau (1 lentelė), kritinis multiplizmas turi būti atskirtas nuo betikslio multiplizmo. Shadish (1986) pritaria minčiai, kad įvertintojai „taiko nepaprastai platų metodų ir savarankiškų teorijų spektrą“ (p. 75). Tačiau šis multiplizmas renkantis „teorijas ir metodus“ nėra susijęs su jokia sąmoningai susiformavusia teorija, kurioje būtų teigiama, kad mokslą geriausiai plėtoti būtų tokiu būdu. Greičiausiai tai buvo sukelta vidinių ir išorinių tam tikros profesijos veiksmų“ (p. 76). Toks daugialypumas yra atsitiktinis, t.y. betikslis užuot buvęs sistemingas arba, kaip sakoma koncepcijos pavadinime, „kritinis“.

Prieštaraudami šitokiam betiksliam multiplizmui, Williams Shadish, Thomas Cook ir šio straipsnio autorius iškelia mintį, jog siekiant efektyviai atlikti tyrimą ir įvertinimą būtina taikyti multiplizmą sistemingai nepaisant kritinių pastabų, išsakytų įvade. Suplanuoto multiplizmo esmė, remiantis Shadish (1986, p.78), „nereiškia, kad kuo daugiau, tuo geriau: daugiau veiksmų, daugiau metodų, daugiau priemonių, daugiau teorijų. Daug svarbiau mokslininkus ap rūpinti priemonėmis ir padėti jiems tirti savo pažinimo ribas.“

Tad „vienas svarbiausių požymių yra heterogeniško nurodymas; svarbus ne taikytų pasirinkimų skaičius, bet labiau tai, kad būtų pasitelkti įvairūs pasirinkimai, kurių *šališkos nuostatos veikia skirtingomis kryptimis*“. Daugialypumas tikslingas tuo atveju, jeigu ir tik jeigu suteikia įvairumo toms problemoms, kurios buvo aptartos įvade ir leidžia atskleisti šališkumą, jau minėtą šio skyriaus pradžioje.

Dėl šios priežasties siekiant gebėjimo taikyti *sisteminį* (arba kritinį) multiplizmą, svarbu išsamiai suprasti mokslo

its problems. It is suggested, then, to use several approaches simultaneously. If the different approaches yield similar results, we can trust that there is something reasonable in them, until there is evidence of the opposite. If they differ considerably, it is obvious that there is a problem: We know that we cannot trust the results, but we do not know which ones are more trustworthy and which ones are less, or in what regard they are trustworthy.

The following are examples of critical multiplism (see also Houts et al., 1986; Patry, 1989; Shadish, 1986; etc.):

- Multiple operationalism and multiple methods (Campbell & Fiske, 1959);
- Analysis of complex multivariate instead of univariate models (this refers to the constructs to be analyzed);
- Checking multiple rival hypotheses (Campbell, 1969);
- Using multiple related studies, as in research programs (Herrmann, 1976), research strategies with several differently designed assessments such as multi-site studies (Seashore Louis, 1984; Herriott & Firestone, 1983; Slavin & Madden, 1993; Audet & d'Amboise, 2001), etc.
- Synthesis of results of independent studies (e.g., through meta-analyses, Cook & Leviton, 1980; Cook, 1991; Glass, McGaw & Smith, 1981; Rosenthal, 1989; etc.);
- Analysis of the same data with different statistical approaches;
- Use of different theories to give hypothetical answers to similar research questions and to interpret results (competition, coexistence or even complementarity of theories, see, e.g., Patry, 1998 and below).

As said above (see also table 1), critical multiplism is to be distinguished from mindless multiplism. Shadish (1986) sustains that evaluators “have used an unusually wide array of methods and substantive theories” (p. 75). However, this “multiplism in theory and method choice mostly has not been due to any conscious theory that science is best conducted in such a fashion. Rather, it probably was elicited by internal and external sociological forces in the profession.” (p.76) This kind of multiplism, hence, is haphazard, i.e., mindless, instead of being systematic or, as the name of the approach suggests, “critical”.

In contrast to this mindless multiplism, authors like William Shadish, Thomas Cook and the present author suggest that in order to achieve the goal of doing valid research and evaluation despite the criticisms mentioned above, it is necessary to apply multiplism systematically. The point of planned multiplism, according to Shadish (1986, p. 78), “is not that more is better – more operations, more methods, more measures, and more theories. Rather, the point is to provide tools to help scientists explore the boundaries of their knowledge.”

And for this, “one of (the) key features is a prescription for heterogeneity; it is not the sheer number of implemented options that are important, but rather that the different options be implemented that *have biases operating in different directions*” (italics added). Multiplism achieves its goals if and only if it provides diver-

kritiką ir mėginti paaiškinti ją remiantis įvairiais požiūriais. Tad čia svarbus:

2 pasiūlymas. Suvokti galimą savo tyrimo šališkumą ir parengti atitinkamas gynybos strategijas pasitelkiant daugialypius modelius!

2 PARADIGMOS IR TEORIJŲ PLIURALIZMAS

Bendras metodologijos principas, kuris taip pat gali būti pritaikytas įvertinimams yra tas, kad net geriausias tyrimo sumanymas ir vertinimo priemonės negali apsaugoti tyrimo rezultatų, jeigu tas tyrimas grindžiamas netinkama teorija. Todėl pirmiausiai norėčiau aptarti pačią teoriją, kuria tyrimas grindžiamas, o po to kalbėti apie vertinimus. Nepaisant visos jų svarbos, teorijos klausimams buvo skiriama daug mažiau dėmesio nei vertinimo klausimams. Tačiau kritiniam multiplizmui teorijų pliuralizmas yra pats svarbiausias klausimas. Galima pridurti, jog šis klausimas svarbesnis įvertinimo tyrimuose, nes teorinis pagrindas šiuose tyrimuose yra labai jautri tema.

2.1. TEORIJOS (ARBA TEORIJŲ) VAIDMUO ĮVERTINIME

Manoma, kad daugelyje įvertinimo tyrimų teorijos nėra svarbios. Iš tiesų tiriamas klausimas įvertinimo studijoje yra ne tai, ar pati teorija teisinga, ar ne (kad ir ką reikštų pasakymas „būti teisingam“, nes, pavyzdžiui, remiantis konstruktyvistų epistemologija, išvis nėra būdų sužinoti tiesą, todėl Glasersfeld (1981) kalba ne apie tiesą, o apie „pasiekiamumą“ – šiame straipsnyje šis klausimas smulkiau nenagrinėjamas). Daug svarbiau, ar konkreti programa veiksminga (pvz., Patry, Perrez, 2000), ar konkreti institucija pasiekia savo tikslus (pvz., Thonhauser, Patry, 1999); iš pirmo žvilgsnio atrodo, kad šiems tikslams teorija nėra būtina.

Tačiau reikia priminti, kad mokslas (įskaitant ir įvertinimą) prasideda nuo klausimų ir problemų kėlimo, po jų priimami preliminarūs sprendimai (Popper, 2002). Tačiau šie preliminarūs sprendimai būtinai yra teoriniai, nesvarbu, ar tai pripažįstama, ar ne. Kalbant toliau, būtina pridurti, kad bet kuris stebėjimas yra susijęs su tam tikra teorija (tai taip pat akcentavo Popper, 2002). Tai taip pat būdinga ir konstruktyvistų epistemologinėje tradicijoje, Glasersfeld, 1981). Tie, kurie įsivaizduoja, jog nesiremia jokia teorija, vis dėlto taiko subjektyvias teorijas arba neprofesionalias teorijas (Furnham, 1998) su visomis iš jų kylančiomis problemomis: jos yra nesistemiškos, dėl „patvirtinimo-paneigimo“ spąstų (žr. aukščiau) dar šališkesnės, pats jų turinys atsitiktinis ir t.t., – bet vis vien tai tam tikro pobūdžio teorijos. Taigi, bet kuriuo atveju geriau vadovautis aiškiais teorijomis, t.y. remtis moksliniais tyrimais, nes bet kuriame socialiniame tyrime vis dar likę daug „pasilėptų“ teorijų.

3 pasiūlymas: prisipažinkite, kad jūs taikote teorijas ir aiškiai jas išdėstykite!

2.2. AIŠKINIMŲ DAUGIALYPUMAS

Reiškiniai paprastai turi ne vieną priežastį; todėl čia tiktų taip vadinamas „INUS“ modelis (*angl.* a cause is “an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition” – priežastis yra „nepa-

sity with respect to the problems addressed in the introduction and hence permits to uncover the biases mentioned at the beginning of the present section.

For this very reason, in order to be able to apply *systematic* (or critical) multiplism, it is essential to be aware of the criticisms to science in detail and to try to account for them through varying approaches. Hence the

suggestion 2: Know the potential biases of your research and work out appropriate defense strategies through multiple approaches!

2 PARADIGMS AND PLURALISM OF THEORIES

A general principle in methodology that holds also for evaluations is that the best research design and assessment tools cannot save a study with a bad theory. Therefore, I want first to look at the theory underlying a study, and then go on with assessments. Despite its importance, issues of theory have much less been looked at than issues of assessment. In critical multiplism, the pluralism of theories is at stake. One might add that this question is the more important in evaluation research because the theoretical background in evaluation studies is a very sensitive topic.

2.1 THE ROLE OF THEORY (OR THEORIES) IN EVALUATION

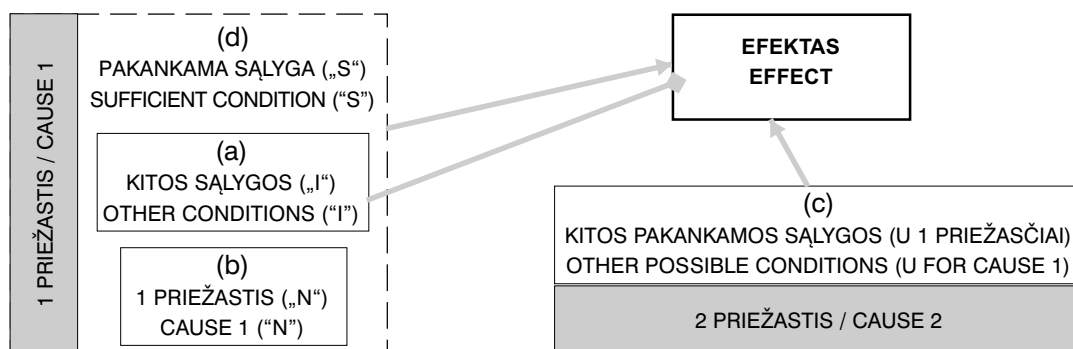
In many evaluation studies, theories supposedly are not relevant. Indeed, the research question in an evaluation study is not whether a theory is true or not (whatever “being true” might mean, since for instance according to constructivist epistemology, there is no means at all to find out truth, and Glasersfeld, 1981, consequently speaks of “viability” instead of truth – it is not the place here to go into details in this regard). Instead, the question is whether a specific program works (e.g., Patry & Perrez, 2000) or a specific institution achieves its aims (e.g., Thonhauser & Patry, 1999); at first sight, for these aims, a theory is not necessary.

However, it is to be reminded that science (including evaluation) starts with questions and problems, and is followed by tentative solutions (Popper, 2002). These tentative solution, however, are necessarily theoretical, whether acknowledged or not. One must add, further, that any observation is also theory-bound (this has also been stressed by Popper, 2002, but also in the constructivist epistemological tradition, Glasersfeld, 1981). Those who pretend not to use any theory actually use subjective theories or lay theories (Furnham, 1988) with all its problems: They are unsystematic, the confirmation-refutation trap (see above) is even more biasing, the content is haphazard, etc. – yet it is still some kind of theory. So in any case it is better to have explicit theories and so to capitalize on scientific research – there is still much implicit theory in any social research left. Hence the

suggestion 3: Acknowledge that you use theories and make them explicit!

2.2. MULTIPLE EXPLANATIONS

Phenomena usually do not have only one cause; rather, the “INUS” concept is appropriate (a cause is “an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition; Mackie, 1974, p. 62, italics in the origi-



1 pav. INUS sąlygos (žr. tekstą)
Figure 1: INUS-conditions (see text)

kankama, tačiau *ne perteklinė* dalis *nebūtinis*, bet *pakankamos sąlygos*“ (Mackie, 1974, p.62).³ INUS principai galėtų būti pavaizduoti šitaip: (1 pav.; kaip pavyzdys naudojamas klasikinis šališkumo socialiniuose tyrimuose atvejis: eksperimentatoriaus laukimo efektas, Rosenthal, 1976 pavyzdyje „Pigmalionas auditorijoje“, Rosenthal, Jacobson, 1968):

(a) *I*. Tai (potenciali) priežastis, bet pati savaime ji *nepakankama*; ypač veiksminga kaip priežastis ji gali tapti tik tada, kai sudaroma grupė kitų sąlygų (1 pav. „kitos sąlygos“). Pigmaliono atveju dėstytojai yra įsitikinę, jog kai kurie jų studentai pasižymi ypatingais gabumais, padedančiais jiems „pražysti“ („atskleisti savo gabumus“), t.y. sparčiai tobulėti. Savaime šie gabumai yra nepakankami, nes be dėstytojo darbo jie nebus veiksmingi.

(b) *N*. Ši (potenciali) priežastis *nėra perteklinė viso* veiksmių „rinkinio“ dalis; tai reiškia, kad „rinkinys“ nebus veiksmingas tol, kol (potenciali) priežastis netaps būtina jo dalimi. „Pigmaliono“ atveju „rinkinys“ apima aukščiau minėtus gabumus ir tinkamą mokymą. Dėstytojai mano, jog ugdyti sėkmingiau sekasi „greitai pražystančius“ („atsiskleidžiančius“), negu kitus studentus.

(c) *U*. Šis „rinkinys“ yra *nebūtinis*; galėtų būti ir kitų panašaus poveikio „rinkinių“, kai nesama arba esama pirmojo „rinkinio“ ar jo elementų, bet nėra (potencialios) priežasties. „Pigmaliono“ atveju galima būtų diskutuoti su Rosenthal ir Jacobson, kad ne studentų gabumai paspartina jų tobulėjimą, bet labiau dėstytojų tikėjimas, verčiantis juos kitaip elgtis su potencialiai greitai „pražystančiais“ („atsiskleidžiančiais“) studentais.

(d) *S*. Šis „rinkinys“ yra *pakankamas*; jeigu turimas visas toks „rinkinys“, rezultatas greičiausiai bus pasiektas, esant bet kokioms sąlygoms. „Pigmaliono“ atveju dėstytojai mano, jog būtent „pražydimo“ (atsiskleidimo) potencialas, bet ne jų tikėjimas ir su tuo susijęs elgesys yra spartėjančio tobulėjimo priežastis.

Iš Pigmaliono pavyzdžio matyti, kad taikant „INUS“ sąlygos modelį, nagrinėjamas gali būti ir šališkumas, pvz., kaip 2 priežastis (1 pav., c komponentas) ir pasiūlyta alternatyvi laukiamo rezultato interpretacija.

nal).³ The INUS principles can be described as follows (see also figure 1; as example, I use a classical bias in social research: the experimenter expectancy effect, Rosenthal, 1976, in the example of “Pygmalion in the classroom”, Rosenthal & Jacobson, 1968):

(a) *I*: There is a (potential) cause, but it is by itself *insufficient*; it can become fully effective as a cause only if a set of other conditions are given (“other conditions” in fig. 1). In the Pygmalion case, the teachers believe that some of their students have a particular aptitude which makes them “bloomers” (they will bloom up quickly), i.e., they will develop particularly fast. This aptitude by itself is insufficient since without teaching it will not become effective.

(b) *N*: This (potential) cause is a *non-redundant* part of the full “package” of influences; this means that the “package” will not become effective unless the (potential) cause is an integral part of it. In “Pygmalion”, the “package” contains the aptitude mentioned above and appropriate teaching. The teachers assume that teaching “bloomers” will be more successful than teaching other students.

(c) *U*: This “package” is *unnecessary*, there might be other “packages” which have a similar influence without the first “package” being present or with the first “package” or elements thereof being present, but not the (potential) cause. In “Pygmalion”, one might argue with Rosenthal and Jacobson that it is not the aptitude of the students which triggers the acceleration in student development, but rather the expectation of the teachers which makes them behave differently towards the potential “bloomers”.

(d) *S*: This “package” is *sufficient*, if the full “package” is provided, the effect is likely to happen, whichever other conditions may occur. The teachers assume in the “Pygmalion” case that indeed it is the blooming potential – and not their expectation and the related behavior – that accounts for the acceleration.

One can see in the Pygmalion example that using the INUS condition model permits to include biases in the explanation, e.g., in the Cause 2 (c component in fig. 1) and to provide an alternative explanation of the presumed effect.

³ Šis teiginys gali būti suformuluotas ir pozityvia nuostata. Tada jis skambėtų taip: „Priežastis yra esminė sąlygų visumos dalis, kurios neišvengiamai veda link pasekmės, bet ši pasekmė taip pat gali turėti ir kitas priežastis, negu tik tą, kurią tyrinėjame“.

³ It could be formulated in a positive manner. It would read then: “A cause in an essential part of package of conditions which inevitably leads to the effect, but this effect can also have other causes than the one under investigation.”

Šis aiškinimo modelis daug sudėtingesnis nei tradicinis aiškinimas, pvz., remiantis Hempel-Oppenheim schema. Bet galvojant apie socialinio elgesio ir veiksmų sudėtingumą, „INUS“ modelis atrodo tinkamesnis, leidžiantis įtraukti daugiau būtinų veiksmų, lemiančių rezultata. Tačiau tie įvairūs veiksniai galėtų būti susieti su įvairiomis teorijomis. Pvz., aukščiau pateiktame pavyzdyje teorija į kurią dėstytojai sutelkia dėmesį, yra gabumų teorija (1 priežastis, „N“ – elementas); tačiau kitos teorijos taip pat svarbios, pavyzdžiui, dėstyto tuotojo poveikio teorija (2 priežastis, „U“ – elementas) ir panašiai. Vadinas, socialinių tyrimų studijoje reikėtų kiek įmanoma išsamiau paaiškinti tam tikro „jeigu tai“ sąryšio „INUS“ modelio sąlygas, kur „jeigu“ komponentas žymi (potencialią) priežastį, o „tai“ komponentas – rezultata. Iš to išplaukia, kad teorinėje pranešimo dalyje įvairios „INUS“ modelio sąlygų dalys turėtų būti kiek įmanoma aiškiau išdėstytos.

Kaip pažymi Shadish, Cook, Campbell (2002, p. 5), „mes retai žinome visus veiksmus ir tai, kaip jie susiję vienas su kitu“. Ypač tai pasakytina apie praktinius tyrimus (kurie yra tipinių tyrimų ir kokybės užtikrinimo įvertinimų pagrindas), kur neįmanoma paaiškinti visų potencialių „INUS“ modelių sąlygojamų veiksmų. Be to, lemiamas veiksnys tai, kad teorijos bet kuriuo atveju yra tikrovės abstrakcijos, kur neatsižvelgiama į kai kuriuos veiksmus, kurie galėtų būti lemiami (taip pat žr. Herrmann, 1979, abstraktumo ir konkretumo dilemą, paminėtą įvade). Nepaisant to, INUS modelis turėtų būti naudingas.

4 pasiūlymas. Užuoat bandžius suprasti po vieną kintamąjį atskirai, reikėtų išanalizuoti sistemą remiantis INUS modeliu, t.y. mėginti nustatyti kiek įmanoma daugiau potencialių INUS veiksmų (ypač b dalyje: rinkinio elementai, kurie nėra dėmesio centre, bei c dalyje: kitas potencialus priežastčių rinkinys)!

3 DAUGIALYPIAI METODAI

Vienas ankstyvųjų kritinio multiplizmo modelių yra daugiasavybė daugiametodė matrica.

3.1. CAMPBELL'O IR FISKE'O DAUGIASAVYBĖ DAUGIAMETODĖ MATRICA

Donald T. Campbellas (1916-1996) gali būti laikomas kritinio multiplicizmo pradininku, nors ir nėra pasiūlęs šio termino, kurį reikėtų sieti su vienu iš Campbello studentų Thomu Cooku (pvz., 1985). Campbellas rašė apie trianguliaciją, o svarbiausiu čia laikoma Campbello ir Fiske'o (1959) darbas apie daugiasavybę daugiametodę matricą (angl. *multitrait-multimethod matrix*). Remiantis šiuo modeliu kiekvienas iš kelių, tikėtina, kad nepriklausomų, konstrukčių (kuriuos autoriai vadina savybėmis; C_1, C_2, C_3 , ir t.t.), vertinami taikant įvairius metodus (M_1, M_2, M_3) ir vis įtraukiant tuos pačius subjektus. Tuomet rezultatus galima sieti tarpusavyje ir išskirti įvairius pagrįstumo (tinkamumo) tipus:

- Konvergencinis pagrįstumas: koreliacija tarp paties konstrukto (C_i) vertinimų taikant skirtingus me-

This approach to explanation is much more complex than the traditional explanation schema, e.g., according to the Hempel-Oppenheim-schema. But looking at the complexity of social behavior and action, the INUS concept seems much more appropriate, giving the possibility to include many factors required for an effect to occur. The different factors, however, may be related to different theories. E.g., in the example above, the theory the teachers are focusing on is a theory of aptitude (Cause 1, “N”-element); however, other theories are relevant as well, such as a theory of teaching (for the other conditions, “I”-element), the theory of experimenter effects (for the Cause 2, “U”-element), and the like. Within a social research study, hence, it would be appropriate to account as fully as possible for the INUS conditions of a particular if-then-relationship, the if-component being the (potential) cause, the then-component being the effect. This means, in the theoretical part of the report, the different parts of the INUS conditions should be made explicit as much as possible.

However, as Shadish, Cook and Campbell (2002, p. 5) have pointed out, “we rarely know all of (the factors) and how they relate to each other”. Particularly in field settings – the typical settings in research and evaluation of quality assurance – it is impossible to account for all potential factors within a set of INUS conditions. This is due, among others, to the fact that theories are necessarily abstractions of reality, neglecting some factors which might be crucial (see also Herrmann's, 1979, generalizability-concreteness dilemma mentioned in the introduction). Nevertheless, it is appropriate to think in INUS terms.

Suggestion 4: Instead of conceiving one independent variable at a time, make an analysis of the system in INUS terms, i.e., try to identify as many potential INUS factors (particularly in the parts b: elements of the package which are not focused, and c: other potential cause package)!

3 MULTIPLE METHODS

One of the earliest approaches used in critical multiplism is the multitrait-multimethod matrix.

3.1. THE MULTITRAIT-MULTIMETHOD MATRIX BY CAMPBELL AND FISKE

Donald T. Campbell (1916-1996) can be seen as the founder of critical multiplism, although he did not coin the term himself which goes back to one of his students, Thomas Cook (e.g., 1985). Instead, Campbell wrote about triangulation, the most important publication in this regard being Campbell and Fiske (1959) on the multitrait-multimethod matrix. According to this approach, each of several presumably independent constructs (which the authors call “traits”; C_1, C_2, C_3 , etc.) are assessed with different types of methods (M_1, M_2, M_3) with the same subjects. One can then correlate the results and distinguish different types of validity:

- Convergent validity: Correlation between assessments of the same construct (C_i) with different methods (M_1, M_2 , etc.); for good assessment tools, this correlation should be as high as possible, the upper limit being in the range of the respective reliabilities.

todus (M_1 , M_2 ir t.t.); ieškant tinkamų vertinimo priemonių ši koreliacija turi būti kiek įmanoma aukštesnė tai aukštesnei reikšmei išliekant atitinkamo patikimumo (tikslumo) ribose.

- diskriminantinis pagrįstumas I: koreliacijos tarp skirtingų konstrukčių (C_1 , C_2 , t.t.) vertinimų taikant tą patį metodą (M_i); ieškant tinkamų vertinimo priemonių ši koreliacija turi būti žema – jei ji aukšta, tai konstruktai nėra nepriklausomi ir/ arba taikant metodą (M_i) gaunami artefaktai.

- diskriminantinis pagrįstumas II: koreliacijos tarp įvairių konstrukčių (C_1 , C_2 , t.t.) vertinimų taikant skirtingus metodus (M_1 , M_2 , etc.); ieškant tinkamų vertinimo priemonių ši koreliacija taip pat turi būti žema, jei ji aukšta (tokia aukšta kaip I diskriminantinio pagrįstumo atveju), tai rodo, kad konstruktai nėra nepriklausomi.

Patry (1989) yra pasiūlęs bendresnę koncepciją, apimančią ne tik vertinimą („priklausomas kintamasis“), bet ir modelį bei nepriklausomą kintamąjį.

Koreliacijos gali būti įtrauktos į koreliacijos matricą, iš čia atsiradęs pavadinimas „daugiasavybė ir daugiametodė matrica“ (sutrumpintai DSDM; *angl.* MTMM). Šis konvergencinio ir diskriminantinio pagrįstumo nustatymo metodas tapo vienu iš veiksmingiausių vertinimo pagrįstumo mechanizmo priemonių. Tačiau atsižvelgiant į daugybę kritinių pastabų, išsakytų įvade, nors tyrimai pagal DSDM dažni, tik keletu atvejų neabejotinai pavyko nustatyti ir konvergentinį, ir diskriminantinį pagrįstumą (Fiske, 1982). Labai dažnai I diskriminantinio pagrįstumo (tas pats metodas, bet skirtingi konstruktai) koreliacijos per aukštos, o tai rodo, kad būtent metodai lemia aukštą kintamumo procentą vertinimuose; kitaip tariant, rezultatus sąlygoja taikomi metodai, o ne tiriamas reiškinys. Tad dar kartą įstikiname, kad kritinis multiplizmas neišsprendžia problemų, bet greičiau atskleidžia (diagnozuoja, nustato) problemą.

5 pasiūlymas. Tarkykime DSDM!

3.2. ŠALIŠKUMAS VERTINIMUOSE

Tačiau diagnozavimas yra būtina sąlyga problemoms spręsti, jeigu toks sprendimas nėra atsitiktinis arba „aklas“ (tai gana įdomi mintis, žr. Campbell, 1960, bet visiškai neekonomiška). Remiantis *1 pasiūlymu*, būtų neteisinga neatsižvelgti į dažną vertinimo šališkumą, tačiau reikėtų paieškoti geresnių metodų.

Ši diagnozė tik parodo, *kad* mūsų vertinimai labai šališki ir atskleidžia tokio šališkumo pobūdį, bet nenurodo, *kodėl* jie tokie šališki. Todėl mums reikalinga tinkama vertinimo teorija.

Egzistuoja keletas teorijų, kurios galėtų būti pasitelktos šiam tikslui. Christensen (1981) yra pasiūlęs „subjekto motyvų paaiškinimą“: vertinimo situacijose subjektai linkę maksimaliai padidinti kitiems žmonėms daromą įspūdį, t.y., jie siekia pozityvaus „savęs pristatymo“ („savirodos“, *angl.* self-presentation). Patry (2004) šią sąvoką išplėtė pritaikęs savo sukurtą situacijos specifiškumo teoriją. Jis pasiūlė algoritmą, kuriuo gali būti pasiremta siekiant nustatyti, ar galima laukti „savęs pristatymo“ šališkumo konkrečioje vertinimo situacijoje, ar ne. Šį algoritmą matome 2 paveiksle.

Svarbiausia čia tai, kad jei subjektas žinos, jog yra stebimas, šis žinojimas veiks jo elgesį (Patry, 1982). Keletas faktorių gali padidinti (arba sumažinti) tikimybę, kad subjektas žino, jog jis yra stebimas. Visų pirma svarbios yra atviros (eksplicitinės) instrukcijos.

- Discriminant validity I: Correlations between assessments of different constructs (C_1 , C_2 , etc.) using the same method (M_i); for good assessment tools, this correlation should be low – if it is high, this indicates that the constructs are not independent and/or the used method (M_i) produces artifacts.

- Discriminant validity II: Correlation between assessments of different constructs (C_1 , C_2 , etc.) using different methods (M_1 , M_2 , etc.); for good assessment tools, this correlation should also be low – if it is high (as high as in discriminant validity I), there is indication that the constructs are not independent.

Patry (1989) has proposed a more general approach, including not only assessment (the “dependent variable”), but also the design and the independent variable.

The correlations can be put in a correlation matrix, hence the name “mutitrait-multimethod matrix” (in short: MTMM). This method of testing both for convergent and discriminant validities has become one of the most powerful tools for validation of assessment instruments. In accordance with the many criticisms mentioned in the introduction above, however, despite frequent studies with MTMM, only few have been really successful in showing both convergent and discriminant validities (Fiske, 1982). Very frequently, the correlations of discriminant validity I (same method, different constructs) are too high, which indicates that the methods account for a high percentage of variability in the assessments; in other words, the results are due to the methods used, not to the phenomenon under investigation. Once again, it turns out that critical multiplism is not the remedy, but rather the diagnosis of the problem.

Suggestion 5: Use MTMM!

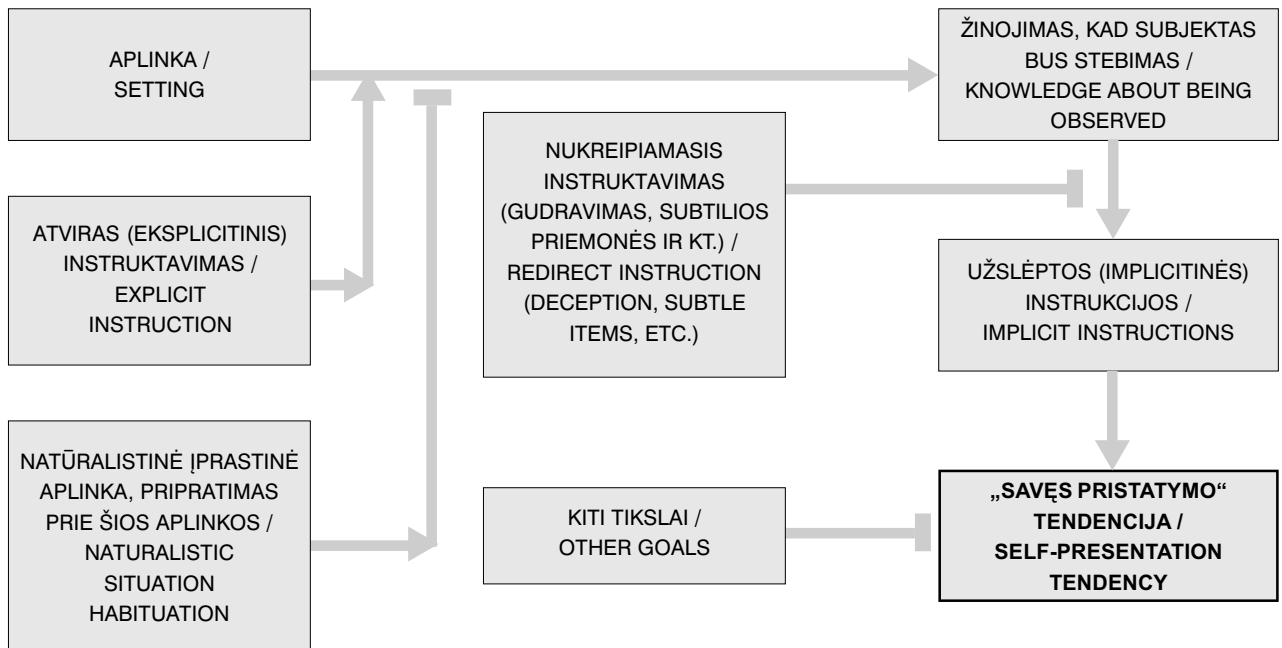
3.2. BIASES IN ASSESSMENT

But the diagnosis is a necessary condition for the remedy if the latter is not to be random or “blind” (which might be an interesting idea, see Campbell, 1960, yet not economic at all). According to *suggestion 1*, it would not be appropriate to close the eyes with respect to the frequent assessment biases, but rather to seek for better methods.

However, this diagnosis is only showing *that* there are strong biases in our assessments (and what kind of biases these might be), but gives no indication as to *why* there are biases. For this we would need an appropriate assessment theory.

There are some theories which can be used for this function. Christensen (1981) has provided a “parsimonious explanation of subject motives”: In assessment situations, subjects tend to attempt to maximize the impression they make upon other people, i.e., they aim at a positive self-presentation. Patry (2004) has extended this concept, using his theory of situation specificity, and has proposed an algorithm which can be used to detect whether one can expect a self-presentation bias in a given assessment situation or not. This algorithm is presented in figure 2.

The most important issue here is that if the subject knows he or she is being observed this will have an impact on his or her behavior (Patry, 1982). Several factors may increase (or decrease) the likelihood that a subject



2 pav. Požymių ir „savęs pristatymo“ (savirodos) tendencijų sąryšis (Patry, 2004, p.46)
 Figure 2: Relationship between the attributes and the self-presentation tendency (from Patry, 2004, p. 46)

Šiose instrukcijose eksperimentuotojas nurodo subjektui, ką jis turi daryti (pvz., atsakyti į klausimyne esančius klausimus, valdyti tam tikrą mechanizmą arba netgi elgtis vienu ar kitu būdu); šitaip eksperimentuotojas apriboja subjekto natūralų elgesį (Tunnell, 1977), o subjektas tegali tik paklusti, jeigu jis tikrai žino apie tokį apribojimą, t.y. supranta, kad yra stebimas.

Aplinka (patalpos ir kitos situacijos charakteristikos, kurios skiriasi nuo natūralios aplinkos; žr. Patry, 1982), kurioje vyksta stebėjimas, taip pat turi įtakos. Jeigu subjektas prašomas eiti į tam tikras patalpas, kur jis bus stebimas, pvz., į eksperimentatoriaus laboratoriją, subjektas negalės nesuvokti, kad bus stebimas. Pakanka tik nežymiai pakeisti aplinką ir subjektas suvoks, kad jis stebimas; pavyzdžiui, tokį poveikį gali turėti patalpoje esanti kamera (tai nustatyta remiantis stebėtų asmenų pastabomis, pvz., Ischi, 1978), net jeigu subjektas ir neturėjo pereiti į kitą vietą dėl tokio tyrimo.

Galima būtų nurodyti ir daugiau veiksnių, tačiau aukščiau paminėti laikytini svarbiausiais. Vis dėlto yra būdų šiems veiksniams apriboti. Pirmiausiai instrukcijas galima nukreipti; viena iš galimybių – pasinaudoti apgaule (priversti žmones patikėti, kad bus stebima kažkas kita, o ne tai, kas tikrai domina; žr. pvz., Miller, 1972). Nors apie etines apgaulės problemas kalbėta (pvz., Baumrind, 1985), tai nėra taip problemiška kaip manyta (Christensen, 1988; P. Patry, 2001). Pagrindinė apgaulės prielaida yra ta, kad žmonėms suteikus klaidingą informaciją apie tyrimo tikslą, jie elgsis tarsi būtų negavę nurodymų, susijusių su svarbiais elgsenos kintamaisiais. Tačiau nors implicitinių instrukcijų kokybė ir kryptis būtų pakeista, tos implicitinės instrukcijos vis vien išliktų. Taigi galima teigti, kad Milgram (1963/1972) paklusnumo eksperimente apgaulė pakeitė implicitines instrukcijas iš „nebūk paklusnus“ (kas būtų spontaniška savęs pristatymo strategija tiems, ku-

knows that he or she is being observed. First of all, the explicit instruction is relevant. This is the instruction an experimenter gives to the subject about what to do (e.g., to respond to a questionnaire, to manipulate some machine or even to behave in a certain manner); by this, the experimenter restricts the subject's natural repertoire of behavior (Tunnell, 1977), and the subject can only comply if he or she is indeed aware that such a restriction is taking place – which means he or she knows about being observed.

A further influence is the setting (the room and other characteristics of the situation which differ from the natural situations; see Patry, 1982) of the observation. If the subject is asked to go to a certain place in order to be observed - e.g., to go to the experimenter's laboratory -, the subject cannot avoid being aware of the observation. It is sufficient to make a slight change in the setting to remind the subject of the fact that he or she is being observed; for instance, the mere presence of a camera in the room may have this effect (as can be deduced from remarks of the observed persons, e.g., Ischi, 1978), even if the subject did not have to move to some other place for the purpose of the study.

One could add more factors, but those seem to be the most important. However, there are some means to influence them. First of all, one can re-direct the instruction; one possibility is deception (making people believe that something else is being observed than what is actually of interest; see, e.g., Miller, 1972). Although ethical problems of deception have been discussed (e.g., Baumrind, 1985), it seems less problematic than usually assumed (Christensen, 1988; P. Patry, 2001). The basic assumption of deception is that people who are given wrong information about the research topic will act as if uninstructed with respect to the relevant behavior variables. However, although the quality and direction of the implicit instruction may have been altered, the implicit instruction itself is not removed. So one could argue that in Milgram's (1963/1972) obedience ex-

rie žino apie Milgram tyrimo tikslą) į „pasirodyk esąs geras subjektas“ (Orne, 1962, prasme) arba „pasirodyk esąs patikimas subjektas“ (Fillenbaum, 1966, prasme) ir panašiai, o tai sudaro dažniausią mokymo(si) eksperimento motyvą. Tačiau subjektas vis vien gali įtarti eksperimentatoriaus ketinimus (McGuire, 1969), tad visa tai implicitiniam instruktavimui suteikia kitokį atspalvį.

Naudojantis klausimynais būtinas tam tikras eksplicitinis instruktavimas; vienas iš būdų „paslėpti“ pradinį tyrimo tikslą – pasitelkti „subtilias“ priemones, t.y. tokias priemones, kurios akivaizdžiai neparodytų kokiam tyrimui jos skirtos (pvz., žr., Timbrook ir kt., 1993). Tai labiausiai siejama su tomis vertinimo priemonėmis, kurias taiko Rorschach – čia nurodymai sąmoningai neapibrėžti (Wallace, 1966), o tai gali sąlygoti žemą šio metodo pagrįstumą (Groth-Marnat, 1990, p. 280).

Socialinėje psichologijoje išrasta daugybė būdų, kaip pakeisti implicitinius nurodymus. Taikant kitus metodus mėginta išvengti implicitinio instruktavimo. Keturi ypač dažnai naudojami būdai apima vadinamąją natūralistinę aplinką, pripratimą prie aplinkos, įtraukimą arba laisvą nevaržomą elgesį (Csikszentmihalyi, 1975) ir elgesio tapimą automatišku arba rutinišku.

Remiantis 2 pav. pavaizduotu modeliu pagrindinis vertinimo šališkumo šaltinis siejamas su subjekto tikslais vertinimo situacijoje. Tai ypač svarbu kai vertinamas įvertinimas, nes socialiniai partneriai įvertindami visuomet turi tam tikrų interesų, kadangi tokios studijos tikslas yra įvertinti esamą padėtį ir numatyti būtinas intervencijas. Šios intervencijos iš tiesų galėtų stipriai paveikti subjektą, kurio elgesys vertinamas. Taigi skirtingai nei tyrimai, kur subjektai nesusiduria su svarbiomis savo elgesio pasekmėmis, įvertinimų metu subjektai labai gerai žino apie vertinimo poveikį, tad tikėtina, jog turi šiuo atžvilgiu savo tikslų.

Šis modelis visiškai suderinamas su „INUS“ koncepcija ir daugialypėmis teorijomis. Pirmiausiai tiriama priežastis yra tam tikra charakteristika, kurią norima įvertinti. Pavyzdžiui, jei ketinama vertinti dėstyto kokybę ir šiuo tikslu atliekama studentų apklausa (pvz., Patry ir Gastager, 2004 panaudotas pavyzdys), tai charakteristikos, kurios bus vertinamos, gali apimti psichologinį klimatą paskaitos metu, dėstytojo geranoriškumą, iššūkių laipsnį dėstant, apmąstymų skaičių auditorijoje, tai, ar tema suprantama, ar ne ir pan. Tai „1 priežastys“ INUS modelyje. Teorijos nuostata yra tokia, kad studentų atsakymai į šiuos klausimus (3 pav.), suteiks žinių apie auditorijoje vyraujančią psichologinį klimatą, t.y. kad atsakymai bus pagrįsti. Bet ji („1 priežastis“) bus veiksminga tik tuo atveju, jei kitos sąlygos patenkinamos, pavyzdžiui, ar studentai perskaito ir supranta klausimą ir pan., o tai jau kitos teorijos klausimas. Be to,

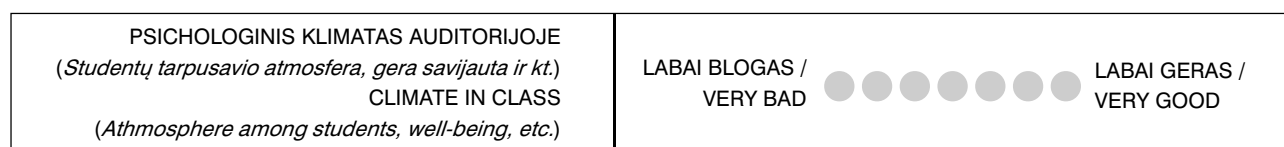
per experiment the deception turned the implicit instruction from “don’t be obedient” (which would be the spontaneous self-presentation strategy for someone who knows about Milgram’s research purpose) to “show that you are a good subject” (in the sense of Orne, 1962) or “show that you are a faithful subject” (in the sense of Fillenbaum, 1966) or the like, which are the likely motives in a learning experiment. Then again, the subject may be suspicious of the experimenter’s intent (McGuire, 1969), which adds another color to the implicit instruction.

The use of questionnaires necessarily requires some explicit instruction; a way to “hide” the initial intention of the study is to use “subtle” items, i.e. items which do not show obviously what they are supposed to measure (see, for instance, Timbrook et al., 1993). This is most extreme in assessment instruments like the Rorschach where the instruction is deliberately ambiguous (Wallace, 1966) - which may have contributed to the low validity (Groth-Marnat, 1990, p. 280f.) of this method.

Social psychologists have been very creative in inventing techniques to alter the subjects’ implicit instruction. Other approaches have been used to avoid an implicit instruction altogether. Four such influence types have been used particularly often: the so-called naturalistic situation, habituation, involvement or flow (Csikszentmihalyi, 1975), and rendering the behavior automatic or routine.

According to the model of figure 2, the central source of assessment bias, hence, is linked with the goals the subject has in the assessment situation. This is particularly important in evaluation assessment because the stakeholders have always specific interests in evaluation situations since the aim of an evaluation study is to value the current state and to draw interventions from this – and these interventions might well have a strong impact on the subject whose behavior is being assessed. So unlike research where the subjects are not faced with important consequences for their behavior, in evaluation, the subjects are very much aware of the impact of the assessment and are likely to have goals in this respect.

This approach is fully compatible with INUS and with the use of multiple theories. First, the cause under investigation is the characteristic one wants to assess. For instance, one wants to assess teaching quality, and for this one questions the students (e.g., as used by Patry & Gastager, 2004). Here the characteristics to be measured are, for instance, the climate in class, the kindness of the teacher, the degree of challenge in teaching, the amount of reflection in class, whether the topic can be understood or not, and the like. These are “Causes 1” in the INUS model. The theory, then, is that the students’ answers to items like the one presented in figure 3 will inform about



3 pav. Klausimyno elemento, skirto vertinti dėstyto kokybę, pavyzdys

(išversta iš Arbeitsgruppe “Evaluation von Lehrveranstaltungen an der Universität Salzburg”, 2001)

Figure 3: Example from an item to assess teaching quality

(translated from Arbeitsgruppe “Evaluation von Lehrveranstaltungen an der Universität Salzburg”, 2001)

studentams turėtų būti suformuluoti tam tikri uždaviniai – pavyzdžiui, kad jie privalo stengtis būti nuoširdūs. Kadangi pateikiamas klausimynas, naudojamos ne tik eksplicitinės, bet ir implicitinės instrukcijos, tad akivaizdu, jog vertinama dėstyto kokybė – šito neįmanoma nuo studentų nuslėpti.

Tačiau studentai gali turėti ir kitų tikslų. Pavyzdžiui, jie gali nenorėti įžeisti dėstytojo pateikdami pernelyg blogus įvertinimus. 1 paveiksle šį tikslą vaizduotų 2 priežastis. Arba „apgaujamą dėstytoją“ (žr. Patry ir Gastager, 2004) priežastis gali turėti įtakos ir studentai pasielgia taip, kaip liepė dėstytojas. Tuomet atsiranda dar vienas nurodymas (tiesioginis arba netiesioginis), kuris gali būti priešingas arba sutapti su pradinio nurodymu („būk nuoširdus“).

DSDM tyrimai parodė, kad metodas (pvz., stebėjimas, o ne klausimynai) turi didelės įtakos vertinimui. I diskriminantinio pagrįstumo koeficientai (įvairios savybės, vienas metodas) turi tendenciją būti aukštesni už tinkamus, o tai rodo, kad nors tiriamos charakteristikos („savybės“) ir skirtingos, rezultatai lieka tokie patys. Tuo tarpu konvergencinio pagrįstumo (viena savybė, skirtingi metodai) koeficientai yra žemesni negu tikimasi: tą pačią charakteristiką („savybę“) vertinant skirtingais metodais, rezultatai labai skiriasi. Tai rodo, kad metodas šališkas.

Campbell (1969) požiūriu, 2 priežastis turėtų būti tikėtina konkuruojanti hipotezė. Apibūdinimas „tikėtina“ teoriškai pagrįstas. Egzistuoja be galo daug įmanomų „2 priežasčių“ ir visų jų išaiškinti neįmanoma. Turėtume atsižvelgti į tikėtinas – ir tai būtų teorija, kuri suteikia žinių apie konkuruojančių hipotezių tikėtinumą. Aukščiau išdėstyta teorija, pasitelkiant „savęs pristatymą“ („savirodą“) ir tikslo struktūrą, yra viena įmanomų teorijų, ji atrodo gana patikima, nes yra patikrinta empiriškai. Tačiau pasitelkus kitas teorijas būtų galima svarstyti kitokias tikėtinas konkuruojančias teorijas (hipotezes).

Šališkumo vertinimuose, pasiremami tikimybe ir aukščiau aptartu modeliu arba algoritmu, o ypač šio modelio apimamų tikslų svarbumu, galėtume pateikti šią formuluotę:

6 pasiūlymas. Vertindami kruopščiai išanalizuokite visus įmanomus tikslus, kuriuos gali turėti subjektai (pvz., „savęs pristatymo“, „savirodos“) ir šių tikslų poveikį vertinamam elgesiui bei pasitelkus daugialypius vertinimus nustatykite galimą šališkumą! Kai kuriais atvejais gali netgi pavykti išvengti šio šališkumo, bet tuomet būtina atidžiai išnagrinėti, ar nekyla kitokio pobūdžio šališkumo pavojų arba kitų problemų (pvz., etinių problemų).

4 APIBENDRINIMAS

Šiame straipsnyje galėjau aptarti tik keletą kritinio multiplizmo taikymo atvejų: daugialypes teorijas ir daugialypius veikimo principus. Svarbiausias kritinio multiplizmo teorijos siūlymas – taikyti apmąstytus daugialypius modelius įvairiuose lygmenyse. Tačiau keletas kliūčių gali kelti pavojų kritinio multiplizmo taikymui.

Teorijų pliuralizmas daugeliui mokslininkų lieka svarbiausia problema. Jei Kuhn (1962) darbą interpretuo-

classroom climate, i.e., that the answers are valid. But this works only if other conditions are satisfied, like the students can read and understand the question and the like – this is matter of another theory. Also, a certain goal structure must be given in the students: For instance, they must try to be honest. Since it is a questionnaire, there is an explicit as well as an implicit instruction, and it is obvious that the assessment deals with teaching quality – there is no way to avoid communicating this to the students.

But the students might have other goals as well. For instance, they might not want to harm the teacher by giving too bad ratings. This goal then would be a Cause 2 in figure 1. Or a “cheating teacher” (see Patry & Gastager, 2004) may try to influence the students, and they want to behave like the teacher asks them to. Here, there is then another instruction (direct or implicit) which may or may not be in opposition to the initial instruction (“be honest”).

In the MTMM research, it has become obvious that the method (e.g., observation as opposed to questionnaires) has a strong impact on the result. Discriminant validity coefficients I (hetero-trait mono-method) tend to be higher than appropriate, which means that even if the characteristics to be assessed (“traits”) are different, the results remain the same, whereas convergent validity (mono-trait hetero-method) are lower than expected: The same characteristic (“trait”), assessed with different methods, yields substantial differences. This would indicate a method bias.

In terms of Campbell (1969), Cause 2 would be a plausible rival hypothesis. The term “plausible” refers, once again, to theory. Indeed, there are infinitely many possible “Causes 2”. It is not possible to account for all of them. One must consider those which are plausible – and it is theory which provides information on the plausibility of rival hypotheses. The theory presented above, using self presentation and the goal structure, is one of the possible theories which seems fairly plausible since it has been empirically tested. However, based on different theories, one might conceive other plausible rival hypotheses.

From the likelihood of biases in assessment, the model or algorithm presented above and, in particular, the importance of the goals within it, we can formulate

suggestion 6: For any assessment, make a careful analysis of possible goals the subjects might have (e.g., of self presentation), of the impact this might have on the assessed behavior, and of possibilities to detect biases through multiple assessments! In some cases it might even be possible to avoid these biases, but then you must carefully analyze if other biases or other problems arouse from this (e.g., ethical problems).

4 DISCUSSION

In the present paper, I could deal only with a few uses of critical multiplism: multiple theories and multiple operationalizations. The suggestion of critical multiplism in general is to use planned multiple approaches on different levels. There are, however, several obstacles which might jeopardize the use of critical multiplism.

Pluralism of theories is a major problem for many scientists. If Kuhn's (1962) work is interpreted as piece of

tume kaip aprašomosios sociologijos mokslo pavyzdį, tai jo aprašoma tai, ką mokslininkai suvokia kaip „gerą tyrimą“, būtent teorijų konkurencija. Šiuo atveju teorijų pliuralizmas laikomas nepriimtiniu. Žinoma, patirtis, įgyta konferencijų metu, skelbiant straipsnius (pvz., recenzuojamuose žurnaluose) ir projektų pasiūlymų apžvalgų metu, rodo, kad socialiniuose moksluose, kaip ir daugelyje kitų mokslinių disciplinų, teorijų gynimas arba jų nuginkijimas yra svarbiausias mokslinės kultūros elementas.

Kaip jau minėta (pvz., Gastager ir Patry, 2000), praktinį darbą dirbantys specialistai, priešingai mokslininkams, sugeba patvirtinti prieštaringas teorijas (pvz., kaip tradicinio ir konstruktivistinio mokymo teorijas) vienu metu. Tačiau šios teorijos nėra tarpusavy susijusios. Prireikus, minėti specialistai gali pagrįsti savo praktinius įgūdžius remdamiesi bet kuria viena teorija (Patry, Schwetz ir Gastager, 2000) ir atmesti kitas. Visa tai kertasi su mokslininkų darbo būdais.

Tačiau nė vienas iš minėtų dviejų būdų – nei abipusė atskirtis, kuria remiasi mokslininkai, nei vienas kuris nors konkretus panaudojimas atsižvelgiant į tam tikrus poreikius, kaip elgiasi praktikai – nėra tinkamas. Kritinio multiplizmo siekiamas teorijų pliuralizmas nereiškia nei teorijų konkurencijos, nei jų koegzistencijos (atskirose „ekologinėse nišose“), bet papildomumą (galimybę jas papildyti). Bent jau turėtų būti pasitelkta „INUS“ koncepcijos modelis, kuriuo remiantis pritaikomi daugialypiai galimi panašaus reiškinio paaiškinimai.

DSDM yra vienas iš būdų visa tai gyvendinti: pasitelkus įvairius vertinimo metodus arba, remiantis Patry (1989), netgi pritaikius skirtingus metodus tiek nepriklausomam, tiek priklausomam kintamajam nustatyti. Tačiau tokie tyrimai brangiai kainuoja, o tai, matyt, yra viena svarbiausių kliūčių taikyti kritinį multiplizmą. Turimi resursai (finansiniai ištekliai, personalas ir kt.) paprastai nėra pakankami naudoti kritinį multiplizmą. Paskutinis,

7 pasiūlymas. Socialinių tyrimų finansavimas turėtų būti gerokai didesnis, kad kritinį multiplizmą būtų galima taikyti. Tai svarbu ir studijų kokybės įvertinimo atveju.

Vadinasi, atsiduriame tarsi užburtame rate: socialinių tyrimų kokybė abejotina (žr. iškeltas problemas įvade), todėl finansavimas menkas. O menkas finansavimas trukdo tyrimų tobulinimą, nors ir žinome, kaip juos tobulinti – pasitelkus kritinį multiplizmą.

(descriptive) sociology of science, it describes what researchers perceive as “good research”, namely competition of theories. In this case, pluralism of theories is seen as inappropriate. Indeed, experience in conferences as well as with publication practices (e.g., with reviewed journals) and project proposal reviews indicates that in social sciences – as in most other scientific disciplines – the defense or challenge of theories is a major element of scientific culture.

We have shown elsewhere (e.g., Gastager & Patry, 2000) that practitioners, in contrast, have no problem in sustaining contradictory theories (like theories of traditional and constructivist teaching) simultaneously; however, these theories are not linked with each other. Upon request, the practitioners can provide practice according to either theory (Patry, Schwetz & Gastager, 2000) and ignore the opposite one. This is in flagrant contrast to the researchers’ practice.

However, neither of the two practices – mutual exclusion, as done by the researchers, or use of either one according to the specific needs, as done by the practitioners – is appropriate. Pluralism of theories as aimed at in critical multiplism means neither competition nor coexistence (in distinct “ecological niches”) of theories, but it means complementarity. At the very least, the “INUS”-structure should be applied, which means that multiple potential explanations of a similar phenomenon should be applied.

The MTMM is one possibility to do so: Using different methods for assessment or, according to Patry (1989), even using different methods both in the independent and in the dependent variable. However, this is very costly. And this is probably one of the most important obstacles to applying critical multiplism: The resources (financial means, personnel, etc.) are usually not such that critical multiplism can be used. Hence the last

suggestion 7: Funding in social research should be much higher so that critical multiplism can be achieved. This holds for evaluation studies as well.

Currently there is a vicious circle in effect: Social Research is of doubtful quality (see the problems in the first section), so there is little funding. And the little funding does not permit the research to improve, although we would know how to improve: through critical multiplism.

LITERATŪRA / REFERENCES

- Amir Y., & Sharon I. (1991). Replication research: A "must" for the scientific advancement of psychology. In Neuliep J. W. (Ed.), *Replication research in the social sciences*. Newbury Park: Sage Publications, p. 51–69.
- Arbeitsgruppe "Evaluation von Lehrveranstaltungen an der Universität Salzburg" (2001) Lehrveranstaltungs-Rückmeldung. Available at <http://www.sbg.ac.at/evaluation/slr-3-vo10sep.doc> Site visited 14th October 2005.
- Audet J., d'Amboise G. (2001). The multi-site study: An innovative research methodology // *The Qualitative Report*, 6(2). Available at <http://www.nova.edu/ssss/QR/QR6-2/audet.html>. Site visited 12th October 2005.
- Baumrind D. (1985). Research using intentional deception // *American Psychologist*, 40, p. 165–174.
- Bornstein R. F. (1991). Publication politics, experimenter bias and the replication process in social science research. In Neuliep J. W. (Ed.), *Replication research in the social sciences*. Newbury Park: Sage Publications, p. 71–81.
- Bungard W., & Bay R. (1982). Feldexperimente in der Sozialpsychologie. In Patry J.-L. (Hrsg.), *Feldforschung. Methoden und Probleme sozialwissenschaftlicher Forschung unter natürlichen Bedingungen*. Bern: Huber, p. 183–205.
- Bunge M. (1967). *Scientific research II: The search for truth*. Berlin: Springer.
- Campbell D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes // *Psychological Review*, 67, p. 380–400.
- 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*, 56, p. 81–105.
- Christensen L. (1981). Positive self-presentation: A parsimonious explanation of subject motives // *The Psychological Record*, 31, p. 553–571.
- Christensen L. (1988). Deception in psychological research: When is its use justified? // *Personality and Social Psychology Bulletin*, 14, p. 664–675.
- Cook T. D. (1985). Postpositivist critical multiplism. In Shortland R. L., & Mark M. M. (Eds.), *Social science and social policy*. Beverly Hills (CA): Sage, p. 458–499.
- Cook T. D. (1991). Meta-analysis: Its potential for causal description and causal explanation within program evaluation. In Albrecht G., Otto H.-U., Karstedt-Henke S., & Bollert K. (Eds.), *Social prevention and the social sciences: Theoretical controversies, research problems and evaluation strategies*. Berlin-New York: Walter de Gruyter.
- Cook T. D., & Leviton L. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis // *Journal of Personality*, 48, p. 449–472.
- Cronbach L. J., & Snow R. E. (1981). *Aptitudes and instructional methods. A handbook for research on interactions*. New York: Irvington.
- Csikszentmihalyi M. (1975). *Beyond boredom and anxiety - the experience of play in work and game*. San Francisco: Jossey-Bass.
- Cziko G. A. (1989). Unpredictability and indeterminism in human behavior: Arguments and implications for educational research // *Educational Researcher*, 18(3), p. 17–25.
- Faust D. (1984). *The limits of scientific reasoning*. Minneapolis: University of Minnesota Press.
- Feyerabend P. K. (1975). *Against method. Outline of an anarchistic theory of knowledge*. London: NLB.
- Fillenbaum S. (1966). Prior deception and subsequent experimental performance: The "faithful" subject // *Journal of Personality and Social Psychology*, 4, p. 532–537.
- Fiske D. W. (1982). Convergent-discriminant validation in measurements and research strategies // *New Directions for Methodology of Social and Behavioral Science*, 12, p. 77–92.
- Foerster H. V. (1998). Entdecken oder Erfinden. Wie läßt sich das Verstehen verstehen? In Carl Friedrich von Siemens Stiftung (Hrsg.), *Einführung in den Konstruktivismus*, 4. Aufl. München: Oldenbourg, p. 41–88.
- Furnham A. F. (1988). *Lay theories. Everyday understanding of problems in the social sciences*. New York: Pergamon Press.
- Gage N. L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press, Columbia University.
- Gastager A., & Patry J.-L. (2000). „Konstruktivistische“ und „traditionelle“ Elemente im Denken von Lehrerinnen und Lehrern (über Handeln im Unterricht). Analyse koexistierender Paradigmen in Subjektiven Theorien rekurrend auf eine Aktualisierung und (Re-) Konstruktion im Dialog-Konsens-Verfahren. Forschungsmitteilungen des Spezialforschungsbereiches F012 der Universität Salzburg „Theorien- und Paradigmenpluralismus in den Wissenschaften – Rivalität, Ausschluß oder Kooperation“, Ausgabe 6.
- Glaserfeld E. von (1981). The concepts of adaptation and viability in a radical constructivist theory of knowledge. In Sigel I., Glinkoff R., & Brodzinsky D. (Eds.), *New directions in Piagetian theory and their application to education*. Hillsdale, NJ. Quoted from Die Begriffe der Anpassung und Viabilität in einer radikal konstruktivistischen Erkenntnistheorie. In Glaserfeld E. von Wissen (1987), *Sprache und Wirklichkeit*. Braunschweig: Vieweg, p. 137–143
- Glaserfeld E. von (1995). *Radical constructivism. A way of knowing and learning*. London: The Falmer Press.
- Glass G. V., McGaw B., & Smith M. L. (1981). *Meta-analysis in social research*. Beverly Hills, Calif.: Sage.
- Groth-Marnat G. (1990). *Handbook of psychological assessment*, second edition. New York: Wiley.
- Hedges L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research // *American Psychologist*, 42, p. 443–455.
- Herriott R. E., & Firestone W. A. (1983). Multisite qualitative policy research: Optimizing description and generalizability // *Educational Researcher*, 12(2), p. 14–19.
- Herrmann T. (1976). *Die Psychologie und ihre Forschungsprogramme*. Göttingen: Hogrefe.
- Herrmann T. (1979). *Psychologie als Problem. Herausforderungen der psychologischen Wissenschaft*. Stuttgart: Klett-Cotta.
- Hetherington J. (1997). Lecture 16: Advanced research design. Available at <http://mccoy.lib.siu.edu/projects/psyc/hetherington/lect16.ppt>. Site visited 5th October 2005.
- Houts A. C., Cook T. D., & Shadish W. R. (1986). The person-situation debate: A critical multiplist perspective // *Journal of Personality*, 54, p. 52–105.
- Ischi N. (1978). Die Erhebung interaktiven Eltern-Kind-Verhaltens durch systematische Beobachtung. In Schneewind K. A., & Lukesch H. (Hrsg.), *Familiäre Sozialisation*. Stuttgart: Klett, p. 27–44.
- Karr C. A., & Larson L. M. (2005). Use of theory-driven research in counseling. Investigating three counseling psychology journals from 1990 to 1999 // *The Counseling Psychologist*, 33, p. 299–326.
- Kuhn T. (1962). *The structure of scientific revolutions*. Princeton, N.J.: Princeton University Press.
- Lewin K. (1951). *Field theory in social science*. New York: Harper & Row.
- Mackie J. L. (1974). *The cement of the universe: A study of causation*. Oxford, England: Oxford University Press.
- Mahoney M. J. (1976). *Scientist as subject: The psychological imperative*. Cambridge, MA: Ballinger.
- McGuire W. J. (1969). Suspiciousness of experimenter's intent. In: Rosenthal, R., & Rosnow, R. L. (Eds.): *Artifact in behavioral research*. New York: Academic Press, 13–57.
- Mertens W., & Fuchs G. (1978). *Krise der Sozialpsychologie. Zur Krisendiskussion über die theoretischen und methodischen Grundlagen der Sozialpsychologie*. München: Ehrenwirth.

- Milgram S. (1963). Behavioral study of obedience // *Journal of Abnormal and Social Psychology*, 67, p. 371–378. Reprinted in Miller A.G. (Ed.) (1972). *The social psychology of psychological research*. New York: The Free Press / London: Collier-Macmillan, p. 82–105.
- Miller A.G. (ed.) (1972). *The social psychology of psychological research*. New York: The Free Press / London: Collier-Macmillan.
- Mischel W. (1968). *Personality and assessment*. New York: Wiley.
- Neuliep J.W., & Crandall R. (1991). Editorial bias against replication research. In Neuliep J. W. (Ed.), *Replication research in the social sciences*. Newbury Park: Sage Publications, p. 85–90.
- Orne M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications // *American Psychologist*, 17, p. 776–783.
- Patry J.-L. (1982). Laborforschung - Feldforschung. In Patry J.-L. (Hrsg.): *Feldforschung. Methoden und Probleme der Sozialwissenschaften unter natürlichen Bedingungen*. Bern: Huber, p. 17–42.
- Patry J.-L. (1989). Evaluationsmethodologie zu Forschungszwecken - Ein Beispiel von „kritischem Multiplizismus“ // *Unterrichtswissenschaft*, 17, p. 359–374.
- Patry J.-L. (1991). Der Geltungsbereich sozialwissenschaftlicher Aussagen: Das Problem der Situationsspezifität // *Zeitschrift für Sozialpsychologie*, 22, p. 223–244.
- Patry J.-L. (1998). Paradigmen - Konkurrenz in der Wissenschaft - Komplementarität im Alltagshandeln? Überlegungen zum Theorie-Praxis-Bezug in der Erziehungswissenschaft. In Schurz G., & Weingartner P. (Hrsg.), *Koexistenz rivalisierender Paradigmen. Eine post-kuhnsche Bestandesaufnahme zur Struktur gegenwärtiger Wissenschaft*. Wiesbaden: Westdeutscher Verlag, p. 87–102.
- Patry J.-L. (2001a). Die Qualitätsdiskussion im konstruktivistischen Unterricht. In Schwetz H., Zeyringer M., & Reiter A. (Hrsg.), *Konstruktives Lernen mit neuen Medien. Beiträge zu einer konstruktivistischen Mediendidaktik*. Innsbruck: StudienVerlag, p. 73–94.
- Patry J.-L. (2001b). Cross-situational consistency of behavior: Triple relevance for research in education // *Salzburger Beiträge zur Erziehungswissenschaft*, 5(1), p. 41–62. Available at http://www.sbg.ac.at/erz/salzburger_beitraege/fruehling2001/jp_2001_1.pdf. Site visited 6th September 2005.
- Patry J.-L. (2004). Situation specificity, validity of the assessment, and the lab-field-problem // *Salzburger Beiträge zur Erziehungswissenschaft*, 8(1), p. 37–52. Available at http://www.sbg.ac.at/erz/salzburger_beitraege/fruehling_2004/patry_1_04.pdf. Site visited 7th September 2005.
- Patry J.-L., & Gastager A. (2004). Dilemmas in quality assessment // *The Quality of Higher Education*, 1. Kaunas: VDU, p. 60–79.
- Patry J.-L., & Perrez M. (2000). Theorie-Praxis-Probleme und die Evaluation von Interventionsprogrammen. In Hager W., Patry J.-L., & Brezing H. (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien: Ein Handbuch*. Bern: Huber, p.19–40.
- Patry J.-L., Schwetz H., & Gastager A. (2000). Wissen und Handeln. Lehrerinnen und Lehrer verändern ihren Mathematikunterricht // *Bildung und Erziehung*, 53, p. 271–86.
- Patry P. (2001). Informed consent and deception in psychological research // *Kriterion*, 14, p. 34–38.
- Popper K. R. (1976). *Logik der Forschung*. Tübingen: Mohr, 6. Auflage (first edition 1934).
- Popper K. R. (2002). *Auf der Suche nach einer besseren Welt*. Vorträge und Aufsätze aus dreißig Jahren, 11. Auflage, Taschenbuchausgabe. München: Piper.
- Rosenthal R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.
- Rosenthal R. (1979). The „file drawer effect“ and tolerance for null results // *Psychological Bulletin*, 86, p. 638–641.
- Rosenthal R. (1989). *Meta-analytic procedures for social research*, 4th print. Newbury Park, CA: Sage.
- Rosenthal R., & Jacobson L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart and Winston.
- Rosenthal R., & Rosnow R. L. (Eds.) (1969). *Artifact in behavioral research*. New York: Academic Press.
- Seashore Louis K. (1984). Multisite/multimethod studies // *Evaluation studies review annual*, 9. Beverly Hills, CA: Sage.
- Shadish W. R. (1986). Planned critical multiplism: Some elaborations // *Behavioral Assessment*, 8, p. 75–103.
- Shadish W. R. (1994). Critical multiplism: A research strategy and its attendant tactics. In Sechrest L. B., & Figueredo A. J. (Eds.), *New directions for program evaluation*. San Francisco: Jossey-Bass, p. 13–57.
- Shadish W. R., Cook T. D. & Campbell D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Slavin R. E., & Madden N. A. (1993). *Multi-site replicated experiments: An application to success for all*. Paper presented at the American Educational Research Association meeting in Atlanta on April 15, 1993.
- Thonhauser J., & Patry J.-L. (Hrsg.) (1999). *Evaluation im Bildungsbereich*. Innsbruck: StudienVerlag.
- Timbrook R. E., Graham J. R., Keiller S. W., & Watts D. (1993). Comparison of the Wiener-Harmon subtle-obvious scales and the standard validity scales in detecting valid and invalid MMPI-2 profiles // *Psychological Assessment*, 5, p. 53–61.
- Tunnell G. B. (1977). Three dimensions of naturalness: An expanded definition of field research // *Psychological Bulletin*, 84, p. 426–437.
- Vygotsky L. (1927). The historical meaning of the crisis in psychology: A methodological investigation. Translated by Van Der Veer R. Available at <http://www.marxists.org/archive/vygotsky/works/crisis/>. Site visited 20th October 2005.
- Wallace J. (1966). An abilities conception of personality: Some implications for personality measurement // *American Psychologist*, 21, p. 132–138.
- Watzlawick P., Beavin J. H., & Jackson D. D. (1974). (4. Auflage) *Menschliche Kommunikation. Formen, Störungen, Paradoxien*. Bern: Huber.
- Webb E. J., Campbell D. T., Schwartz R. D., Sechrest L., & Grove J. B. (1981). *Nonreactive measures in the social sciences*. Boston: Houghton Mifflin (second edition).

*Įteikta 2005 m. spalio mėn.
Delivered 2005 October*

JEAN-LUC PATRY

Zalcburgo universiteto švietimo tyrimų katedros profesorius.
Mokslinių interesų kryptys:
situacijų specifiškumas, švietimo vertybės, įvertinimas,
teorijos ir praktikos ryšys, konstruktivizmas.

Zalcburgo universitetas
Švietimo tyrimų katedra

Akademiestrasse 26, A-5020 Salzburg, Austria
jean-luc.patry@sbg.ac.at

JEAN-LUC PATRY

Professor at the Department of Education
Research interests:
situation specificity, values education, evaluation, theory-
practice relationship, constructivism, etc.

University of Salzburg
Department of Education