Year: 2012

Types of research output profiles: a multilevel latent class analysis of the Austrian Science Fund's final project report data

Mutz, Rüdiger ; Bornmann, Lutz ; Daniel, Hans-Dieter

Abstract: Starting out from a broad concept of research output, this article looks at the question as to what research outputs can typically be expected from certain disciplines. Based on a secondary analysis of data from final project reports (ex post research evaluation) at the Austrian Science Fund (FWF), Austria's central funding organization for basic research, the goals are (1) to find, across all scientific disciplines, types of funded research projects with similar research output profiles; and (2) to classify the scientific disciplines in homogeneous segments bottom-up according to the frequency distribution of these research output profiles. The data comprised 1,742 completed, FWF-funded research projects across 22 scientific disciplines. The multilevel latent class (LC) analysis produced four LCs or types of research output profiles: 'Not Book', 'Book and Non-Reviewed Journal Article', 'Multiple Outputs', and 'Journal Article, Conference Contribution, and Career Development'. The class membership can be predicted by three covariates: project duration, requested grant sum, and project head's age. In addition, five segments of disciplines can be distinguished: 'Life Sciences and Medicine', 'Social Sciences/Arts and Humanities', 'Formal Sciences', 'Technical Sciences', and 'Physical Sciences'. In 'Social Sciences/Arts and Humanities' almost all projects are of the type 'Book and Non-Reviewed Journal Article', but, vice versa, not all projects of the 'Book and Non-reviewed Journal Article' type are in the 'Social Sciences/Arts and Humanities' segment. The research projects differ not only qualitatively in their output profile; they also differ quantitatively, so that projects can be ranked according to amount of output

Types of research output profiles: A multilevel latent class analysis of the Austrian Science Fund's final project report data

Rüdiger Mutz[1,*], Lutz Bornmann[2] and Hans-Dieter Daniel[1,3]

[1]*Professorship for Social Psychology and Research on Higher Education, ETH Zurich, D-GESS, Muehlegasse 21, 8001 Zurich, Switzerland,* [2]*Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, 80539 Munich, Germany and* [3]*Evaluation Office, University of Zurich, 8001 Zurich, Switzerland*
*Corresponding author. Email: mutz@gess.ethz.ch

Starting out from a broad concept of research output, this article looks at the question as to what research outputs can typically be expected from certain disciplines. Based on a secondary analysis of data from final project reports (ex post research evaluation) at the Austrian Science Fund (FWF), Austria's central funding organization for basic research, the goals are (1) to find, across all scientific disciplines, types of funded research projects with similar research output profiles; and (2) to classify the scientific disciplines in homogeneous segments bottom-up according to the frequency distribution of these research output profiles. The data comprised 1,742 completed, FWF-funded research projects across 22 scientific disciplines. The multilevel latent class (LC) analysis produced four LCs or types of research output profiles: 'Not Book', 'Book and Non-Reviewed Journal Article', 'Multiple Outputs', and 'Journal Article, Conference Contribution, and Career Development'. The class membership can be predicted by three covariates: project duration, requested grant sum, and project head's age. In addition, five segments of disciplines can be distinguished: 'Life Sciences and Medicine', 'Social Sciences/Arts and Humanities', 'Formal Sciences', 'Technical Sciences', and 'Physical Sciences'. In 'Social Sciences/Arts and Humanities' almost all projects are of the type 'Book and Non-Reviewed Journal Article', but, vice versa, not all projects of the 'Book and Non-reviewed Journal Article' type are in the 'Social Sciences/Arts and Humanities' segment. The research projects differ not only qualitatively in their output profile; they also differ quantitatively, so that projects can be ranked according to amount of output.

Keywords: research funding; multilevel latent class analysis; research output profiles; ex post evaluation.

1. Introduction

Research funding organizations have shown increasing interest in ex post research evaluation of the funded projects (European Science Foundation 2011a). For instance, the Austrian Science Fund (FWF), Austria's central funding organization for the promotion of basic research and the subject of this article, has conducted ex post research evaluations for some years now (Dinges 2005). By collecting and analysing information on the 'progress, productivity, and quality' (European Science

Foundation 2011b: 3) of funded projects, research funding organizations hope 'to be able to identify gaps and opportunities, avoid duplication, encourage collaboration, and strengthen the case for research' (European Science Foundation 2011b: 3). As stated succinctly in the title of a 2011 working document by the European Science Foundation (ESF), a central topic in this connection is 'The Capture and Analysis of Research Outputs' (European Science Foundation 2011a). This involves the issues of what research outputs are actually important for ex post research evaluation, how they can be classified (typology) and how the

data can be analysed. The ESF document provides the following definition of outputs: 'Research outputs, as the products generated from research, include the means of evidencing, interpreting, and disseminating the findings of a research study' (European Science Foundation 2011a: 5).

But opinions differ on what research output categories should be included in ex post research evaluation. Without doubt, publication in a scientific journal is viewed in all scientific disciplines as the primary communication form (European Commission 2010). For assessing the merits of a publication, bibliometric analyses are favoured. In the humanities and social sciences, however, the use of classical bibliometric analysis (Glänzel 1996; Nederhof et al. 1989; Nederhof 2006; Van Leeuwen 2006) is viewed critically in the face of different forms of research outputs (e.g. monographs) and limitations of the databases (Cronin and La Barre 2004; Hicks 2004; Archambault et al. 2006). For these disciplines, other forms of quantitative evaluation are under discussion (Kousha and Thelwell 2009; White et al. 2009).

A number of authors have made a plea for extending classical biblio analysis and for broadening the concept of 'research output' generally (Bourke and Butler 1996; Lewison 2003; Butler 2008; Huang and Chang 2008; Linmans 2010; Sarli et al. 2010): 'A fair and just research evaluation should take into account the diversity of research output across disciplines and include all major forms of research publications' (Huang and Chang 2008: 2018). Huang and Chang (2008) looked at an empirical analysis conducted of the publication types of all publications in the year 1998–9 across all disciplines at the University of Hong Kong and found that journal articles accounted for 90% and 99% of the total publications produced only in the disciplines medicine and physics. The other disciplines produced output in the form of very different types of written communication, such as books, book chapters, and conference and working papers. Huang and Chang's (2008) comprehensive review of the literature on the characteristics of research output showed that especially in the humanities and social sciences, books, monographs, and book chapters are important forms of written communication.

The German Research Foundation (DFG), Germany's central funding organization for basic research, carried out a survey in the year 2004 on the publishing strategies of researchers with regard to open access (Deutsche Forschungsgemeinschaft 2005), and 1,083 DFG-funded researchers responded (response rate of 67.7%). When the researchers were asked to name their preferred form of traditional publication of their own work, they mentioned articles in scientific journals (on the average about 20 articles in 5 years). Life scientists published the largest number of journal articles (23.6 articles in 5 years) and humanities scholars and social scientists the fewest (12.7 articles in 5 years). Papers in proceedings were published far more often by engineering scholars than by researchers in other

disciplines. Social scientists and humanities scholars had a greater preference for publishing their work in edited volumes and monographs than researchers in other disciplines. However, big differences in the numbers reported (e.g. number of books, number of journal articles) were found within disciplines. This study and the Huang and Chang study made it clear that not only the sciences and humanities differ greatly from other disciplines in their preferred form of written communication. There are great differences also within the natural sciences and humanities. The Expert Group on Assessment of University-Based Research set up by the European Commission came to similar conclusions (European Commission 2010: 26). In the opinion of the expert group, the peer-reviewed journal article is used as the primary form of written communication in all scientific disciplines. In addition, engineering scientists primarily publish in conference proceedings, whereas social scientists and humanists show a wide range of research outputs, with monographs and books as the most important forms of written communications.

The broadest concept of research output is used by the Research Council UK (RCUK) (see www.rcuk.ac.uk), the United Kingdom's (UK) central funding organization, and the Research Assessment Exercise (RAE) (www.rae.ac.uk), which in 2014 will be replaced by the new system, Research Excellence Framework (REF) (ww.ref.ac.uk). RAE and REF have the task of assessing the quality of research in higher education institutions in the UK. Whereas the RAE focuses on scientific impact, the performance measurement by the REF in addition includes societal impact—that is, any social, economic or cultural impact, or benefit beyond academia. As research output, the RAE and REF include different forms of research products (journal article, book, conference contribution, patent, software, Internet publication, and so on). The Research Outcome System (ROS) of RCUK distinguishes a total of nine categories of research outputs: publication, other research output, collaboration, communication, exploitation, recognition, staff development, further funding, and impact. The new REF is planned to extend the currently peer-supported RAE with a quantitative, indicator-based evaluation system that includes bibliometric and other quantitative methods. Butler and McAllister (Butler and McAllister 2009, 2011) spoke generally of a metric as opposed to peer review that would capture more than the classical bibliometric analysis based on journal articles does. RAE and REF are based on a research production model (Bence and Oppenheim 2005) that differentiates between inputs (personnel, equipment, overheads), research generation processes, outputs (paper, articles, and so on), and utilization of research (scientific and societal impact). This kind of structuring in input, process, output, outcome/impact is also found in other frameworks for research evaluation, such as in the payback approach (Buxton and Haney 1998; European Commission 2010; Banzi et al. 2011) and other national and international evaluation systems (European Commission 2010).

2. Limitations of previous research, goals, and research questions

Previous research on research outputs has had the following limitations:

(1) As the databases for the empirical analysis, studies up to now used mainly literature databases (Glänzel 1996; Nederhof et al. 1989) and (survey) data from researchers (Deutsche Forschungsgemeinschaft 2005; Huang and Chang 2008). Therefore, the unit of analysis was people and not projects (European Science Foundation 2011). But the different research outputs and also inputs (e.g. human resources, funding) are tied with the research projects.

(2) For the individual disciplines, the frequencies of certain research outputs were presented mostly in totals and separately without any closer examination of the combination of different research outputs in the form of a core profile. For example, some disciplines focus more on monographs and conference contributions and not so much on journal articles, whereas for other disciplines it is just the opposite. Beyond that, the variability of research output within a discipline, such as that found in a study conducted by the DFG (Deutsche Forschungsgemeinschaft 2005), was hardly considered.

(3) The studies often did not describe the research output comprehensively, as the RAE, REF, and RCUK do, for instance, and instead restricted the study to a specific research output category, such as journal articles. This can lead to an inadequate treatment of some disciplines. Technical sciences can be at a disadvantage, for instance, if patents are not included in the study. Moreover, mostly only selected disciplines were included in the analyses, such as social sciences and humanities, so that comparative analysis of various disciplines was not possible. But research projects in different disciplines can be very similar in the profiles of research output categories (abbreviated in the following as 'research output profiles').

(4) The studies did not distinguish between quality and quantity of research outputs. For example, life sciences are similar to natural sciences in research output profiles, but life sciences have a higher volume of journal articles than the natural sciences do (Deutsche Forschungsgemeinschaft 2005).

The goals of our study are:

Based on a secondary analysis of data in final project reports (Glass 1976) at the FWF, Austria's central funding organization for basic research, the goals of this study were (1) to find, across all scientific disciplines, types of funded research projects with similar research output profiles; and (2) to classify the scientific disciplines in homogeneous segments (e.g. humanities, natural sciences, engineering sciences) bottom-up according to the frequency distribution of these research output profiles. We aimed to establish the types of funded research projects using multilevel latent class analysis (MLLCA) (Vermunt 2003; Kimberly and Muthén 2010; Mutz and Seeling 2010; Mutz and Daniel 2012).

The research questions are:

(1) Are there any types of FWF-funded projects that have different core profiles of research outputs?
(2) Do types of research output profiles vary across scientific disciplines? Can disciplines be clustered into segments according to the different proportions of certain types of research output profiles?
(3) How does the probability of being in a particular type of research output profile depend on a set of project-related covariates (e.g. requested grant sum)?
(4) Is there any additional variability within types of research output profiles that allows for a quantitative ranking of projects according to higher or lower research productivity?

3. The Austrian Science Fund

The FWF is Austria's central funding organization for the promotion of basic research. It is equally committed to all scientific disciplines. The body responsible for funding decisions at the FWF is the board of trustees, made up of 26 elected reporters and 26 alternates (Bornmann 2012; Fischer and Reckling 2010; Mutz, Bornmann and Daniel 2012a, 2012b; Sturn and Novak 2012). For each grant application, the FWF obtains at least two international expert reviews (ex ante evaluation). The number of reviewers depends on the amount of funding requested. The expert review consists (among other things) of an extensive written comment and a rating providing an overall numerical assessment of the application. At the FWF board's decision meetings, the reporters present the written reviews and ratings of each grant application. In the period from 1999 to 2009 the approval rate of proposals was 44.2%. Since 2003, all funded projects are evaluated after completion (Dinges 2005) (see www.fwf.ac.at/de/projects/evaluation-fwf.html). The FWF surveys the FWF-funded researchers, asking them to report the outputs of their research projects using a category system that is akin to the research output system of RCUK. Additionally, referees are requested to provide a brief review giving their opinions on aspects of the final project report. They are also requested to assign a numerical rating to each aspect. The final reports were used for accountability purposes and to improve the quality of FWF's decision procedure (Dinges 2005).

4. Methods

4.1 Data

The data for this study comprised 1,742 FWF-funded research projects called 'Stand-Alone Projects' across all

fields of science (22 scientific disciplines classified into six research areas), which contributed to 60% of all FWF grants ('Stand-Alone Projects', 'Special Research Programs', 'Awards and Prizes', 'Transnational Funding Activities') and finished within a period of 9 years (2002–10). The labelling of the scientific disciplines and the research areas was adopted from the FWF (Fischer and Reckling 2010). Each project head was requested to report the results of his or her research project by completing a form (final project report) containing several sections (summary for public relations; brief project report; information on project participants; attachments; collaboration with FWF).

Of the 1,742 completed FWF-funded research projects (Table 1), most were in the natural sciences (31.6%), and the fewest were in the social sciences (6.0%) and technical sciences (4.5%). The finished projects (end of funding) were approved for funding in the period 1999–2010, one-third of them in 2003–4 alone. Due to still ongoing research projects, projects approved for funding in 2007–8 make up only 3.9% of the total database of 1,742 FWF-funded research projects. The average duration of the research projects was 39 months. In 84.5% of the projects, the project heads were men. The average age of the project heads was 47.

The following six research output categories were captured in quantity and number (count data) and served as the basis for the analysis: publication (peer-reviewed journal article; non-peer-reviewed journal article, monograph, anthology, mass communication, i.e. any kind of publication in mass media, e.g. newspaper article), conference contribution (invited paper, paper, poster), award, patent, career development (diploma/degree, PhD dissertation, habilitation thesis) follow-up project (FWF funded or not). It was not differentiated between different sub-categories of the mentioned research output categories. For example, hybrid, open access and standard peer-reviewed journal articles or ongoing or terminated PhD dissertations were summarized under the respective research output category. In order to avoid problems with different publication lags, the FWF treated equally manuscripts, already published, and manuscripts, accepted for publication. The ex post evaluation approach of the FWF does not distinguish between project publications written in English and written in any other language.

Because of strongly skewed distributions, the count variables were transformed in 2-point to 5-point ordinal scale variables with at most equally sized ordinal classes, to avoid sparse classes or cells in a multivariate statistical analysis. To draw up a typology, actually, binary variables might be sufficient in which it was coded whether the particular research output category (e.g. monograph) existed (= 1) for a research project or not (= 0). However, because we wanted to differentiate a qualitative dimension (types) and a quantitative dimension (amount of output),

we chose an ordinal scale with a sparse number of ordinal classes that in addition allow a quantitative assessment.

The research output variables (Table 2) show a large share of zeros. The most frequently produced types of publication were reviewed journal articles (an average of five per project) and conference papers (on average nine), with a large variance across the research projects. For publication of research results, monographs are used the least (0.2 monographs per project).

In a review of the literature Gonzalez-Brambila and Velosos (2007) discuss age, sex, education, and cohort effects as empirically investigated determinants of research outputs. In our study, we included the following covariates to predict research profile type membership (Table 1): time period of the approval decision, time period of the project end, project duration; overall rating of the proposal, requested grant sum; gender and age of the project head. This information was taken from an ex ante evaluation of the project proposals. In the ex ante evaluation, two to three reviewers rated each proposal on a scale from 1 to 100 (ascending from poor to excellent). The mean of the overall ratings of a proposal averaged across reviewers was 89.7 (minimum: 61.7, maximum: 100).

4.2 Statistical procedure

Latent Class Analysis (LCA) in its basic structure can be defined as a statistical procedure that extracts clusters of units (latent classes (LCs)) that are homogenous with respect to the observed nominal or ordinal scale variables (McCutcheon 1987). Similar to factor analysis, LCs are extracted in such a way that the correlations between the observed variables should vanish completely within each LC (local stochastic independence). LCA is favoured towards cluster analysis due to the fact that fewer pre-decisions are required than in common cluster analysis procedures (e.g. similarity measure, aggregation algorithm). Efficient algorithms for parameter estimation (maximum likelihood) are used, and a broad range of different models (LCA, IRT models, multilevel models, and more) are offered (Magidson and Vermunt 2004; Vermunt and Magidson 2005a). In a more advanced version of LCA, MLLCA, the nested data structure is additionally considered. In our study, research projects are nested within certain scientific disciplines; LCs or project types might vary between scientific disciplines. In MLLCA, not only are projects grouped according to their output profiles but also scientific disciplines will be segmented according to their different proportions of types of output profiles. In the technical framework of MLLCA, LCs represent the types of research output profile, and latent clusters (GClass) indicate the segments of disciplines. It will be presumed that a project in a certain LC behaves the same way (same research output profile) irrespective of the latent cluster to which the project belongs.

Table 1. Sample description (N = 1,742 completed FWF-funded research projects)

Variable	N	Per cent	M	SD	Range
Research area					
Biosciences	399	22.9			
Humanities	339	19.5			
Human medicine	269	15.4			
Natural sciences	551	31.6			
Social sciences	105	6.0			
Technical sciences	79	4.5			
Time period of the approval decision					
1999–2000	210	12.1			
2001–2	433	24.9			
2003–4	582	33.4			
2005–6	448	25.7			
2007–8	69	3.9			
Time period of the project end					
2002–4	281	16.1			
2005–6	531	30.5			
2007–8	558	32.0			
2009–10	372	21.4			
Project duration [months]	1,742	100.0	39.0	8.8	9→62
Overall rating of the proposal (ex ante evaluation)	1,735	99.6	89.7	4.7	61.7→100
Requested grant sum [1,000 €]	1,742	100.0	179.7	82.8	7.6→592.7
Project head's sex					
Man (=0)	1,472	84.5			
Woman (=1)	270	15.5			
Project head's age	1,739	99.8	47.1	9.8	27→87

Note: N = frequency, per cent = column per cent, M = mean, SD = standard deviation, range = minimum and maximum.

In secondary analysis the problem frequently arises that the assumption of local stochastic independence does not fully hold. For instance, career development output categories like diploma/degree and PhD dissertation are more strongly correlated with one another than with the other research output categories, so that a LCA cannot completely clarify the association between the two career development outputs. There are three possible ways to handle this problem (Magidson and Vermunt 2004): First, one or more direct effects can be added that account for the residual correlations between the observed research output variables that are responsible for the violation of the local stochastic independence assumption. Second, one or more variables that are responsible for high residual correlations can be eliminated. Third, the number of latent variables (LCs, continuous latent variables) is increased. In this study we used all three strategies. After a first model run, the residuals were inspected, and a few direct effects were included in the MLLCA model. Additionally, two variables that were responsible for high residual correlations were eliminated—non-peer-reviewed journal articles and diplomas/degrees. Last but not least a MLLCA model was tested that incorporates a continuous latent variable comparable to a factor analysis. With this C-factor not only can residual correlations among the output variables be explained but also additional quantitative differences

between research projects (amount of research output) can be assessed and can be taken for a ranking of projects, respectively. If, over and above, a model fits the data with the *same structure* (i.e. loadings of the research output variables on the factor) for all LCs as well as or better than a model with *different structures* in terms of different loadings of the variables in each LC, all research projects can be compared or ranked on the same scale of the latent variable.

For statistical analysis of the data we used MLLCA as implemented in the software program Latent GOLD 4.5 (Vermunt and Magidson 2005b). Following Bijmolt, Paas, and Vermunt (2004), Lukočienė, Varriale, and Vermunt (2010) and Rindskopf (2006), in a *first step* we calculated a simple LCA of the research outputs to obtain types of research projects with a similar research output profile. To determine the optimal number of classes (project types, segments of disciplines), information criteria were used, such as the Bayesian information criterion (BIC) or Akaike information criterion (AIC). The lower BIC or AIC the better the model fits. These information criteria penalize models for complexity (number of parameters), making it possible to make direct comparisons among models of different numbers of parameters. Results of a simulation study conducted by Lukočienė and Vermunt (2010) for MLLCA models showed that in all simulation conditions, the more advanced criteria AIC3 (Bozdagon

Table 2. Data description (N = 1,742 FWF-funded research projects)

Research output	Scale	Ordinal categories					M	SD	Max	R^2
		0	1	2	3	4				
Journal article, reviewed	Number	0	1–2	3–6	>6		5.1	6.9	75	0.61
	Per cent	23.7	22.8	26.4	27.1					
Journal article, non-reviewed	Number	0	1	2–4	>4		2.8	5.6	50	–
	Per cent	0.50	0.14	0.18	0.18					
Contribution to anthologies	Number	0	1	>1			0.8	2.3	32	0.15
	Per cent	75.4	10.2	14.4						
Monograph	Number	0	>0				0.2	0.7		0.15
	Per cent	89.4	10.6						8	
Mass communication	Number	0	1	>1			1.0	2.9	38	0.16
	Per cent	68.5	13.5	17.9						
Award	Number	0	1	>1			0.5	1.2	13	0.28
	Per cent	74.0	13.5	12.5						
Other output (patent, impact)	Number	0	1	>1			0.6	1.4	26	0.19
	Per cent	71.0	14.9	14.1						
Conference paper	Number	0	1	3–5	6–11	>11	9.1	11.1	101	0.59
	Per cent	12.7	14.9	21.8	24.8	25.8				
Other conference contribution	Number	0	1–2	3–6	>6		4.7	7.5	98	0.51
	Per cent	31.6	20.3	23.9	24.2					
Habilitation thesis	Number	0	1	>1			0.6	0.9	7	0.12
	Per cent	60.7	25.8	13.5						
PhD dissertation	Number	0	1	2	>2		1.1	1.4	23	0.30
	Per cent	41.0	30.8	17.3	10.9					
Diploma/degree	Number	0	1	2	>2		1.3	2.1	22	–
	Per cent	53.4	17.2	10.8	18.6					
Follow-up project	Number	0	1	>1			0.7	1.1	15	0.19
	Per cent	61.6	23.1	15.3						

Note: Per cent = row per cent, M = mean of the raw data, SD = standard deviation of the raw data, Max = maximum, R^2 indicates how well an indicator is explained by the final LC model.

1993) and the BIC(k) outperformed the usual BIC to identify the true number of higher-level LCs (Lukočienė, Varriale and Vermunt 2010). Unlike BIC, BIC(k) uses the number of groups, here the number of disciplines, in the formula for sample size n: BIC(k) $= -2 * \text{LL} - \text{df} * \ln(k)$; AIC3 $= -2 * \text{LL} - 3 * \text{df}$, where df denotes the degrees of freedom, LL denotes the loglikelihood. In the *second step*, we took the hierarchical structure of data into account, calculating an MLLCA to obtain latent clusters of scientific disciplines, or segments. In a *third step* we fixed the number of latent clusters of the second step and again determined the number of LCs. However, Lukočienė and Vermunt's (2010) simulation study showed that the third step results in very small improvement of 1%. We therefore abstained from applying this step.

In the last step we included covariates in the model to explain the LC membership (Vermunt 2010). However, this one-step procedure has the disadvantage that by including the covariates, the model and its parameters, respectively, could change. Therefore, a three-step procedure was suggested. First, we estimated a LC model. Second, we assigned the subjects to the LCs according to their highest posterior class membership probability. Third, the LCs were regressed on a set of covariates using a multinomial

regression model. However, this procedure does not take into account the uncertainty of class membership. Bolck, Croon, and Hagenaars (2004) showed that such a modelling strategy underestimates the true relationships between LCs and covariates. Recently, Vermunt (2010) developed a procedure that takes into account the uncertainty of class membership by including the classification table that cross-tabulates modal and probabilistic class assignment (Vermunt and Magidson (2005b) as weighting matrix into the multinomial regression model. We followed this improved three-step approach. The covariates mentioned above were included for prediction of class membership (Table 1).

5. Results

5.1 Latent structure of research output profiles

In the first step the nested data structure (projects are nested within scientific disciplines) was ignored, and simple LC models were explored. Table 3 shows the results of fitting the models containing one to 11 LCs with and without a continuous latent C-factor, respectively. For model comparison we used the AIC3. Out of all 22 models, Model 15 with four LCs, 107 parameters,

Table 3. Fit statistics for exploratory LC models (project types)

		Without C-factor				With C-factor		
MNR	NCL	LL	NPAR	AIC3	MNR	LL	NPAR	AIC3
1	1	−18,956.3	27	37,993.5	12	−18,136.8	38	36,387.6
2	2	−18,235.5	39	36,588.0	13	−17,938.1	61	36,059.3
3	3	−18,116.0	51	36,386.0	14	−17,836.3	84	35,924.5
4	4	−17,970.9	63	36,130.7	15	−17,785.1	107	35,891.2
5	5	−17,911.4	75	36,047.9	16	−17,758.6	130	35,907.1
6	6	−17,876.8	87	36,014.5	17	−17,728.8	153	35,916.6
7	7	−17,844.5	99	35,985.9	18	−17,695.1	176	35,918.1
8	8	−17,817.8	111	35,968.5	19	−17,679.8	199	35,956.5
9	9	−17,792.1	123	35,953.2	20	−17,638.8	222	35,943.7
10	10	−17,772.2	135	35,949.3	21	−17,610.1	245	35,955.2
11	11	−17,755.5	147	35,951.9	22	−17,592.7	268	35,989.3

Note: MNR = model number, NCL = number of latent classes, LL = loglikelihood, NPAR = number of parameter, AIC3 = Akaike information criterion 3. Final model grey coloured.

and one C-factor shows the smallest AIC3. We therefore decided on this model. With regard to our research questions, there were four types of projects with different research output profiles (qualitative dimension). Additionally, the projects differed in their productivity, i.e. the amount of outputs, represented by the continuous latent C-factor (quantitative dimension).

Figure 1 shows the four LCs or project types with different research output profiles. The 2-point to 5-point ordinal scales were re-scaled such that the numerical values varied within the range of 0–1.0 (Vermunt and Magidson 2005b: 117). We obtained this scaling by subtracting the lowest observed value from the class-specific mean and dividing the results by the range, where the range was nothing but the difference between highest and lowest value. The advantage of this scaling is that all variables can be depicted on the same scale as the class-specific probabilities for nominal variables. It must be noted that the LC results depicted in Fig. 1 were the results of the final MLLCA model (introduced in Section 5.2) and not the non-nested LC model in Table 3. However, this does not matter, because the LC models with and without nesting do not differ.

The four LCs or project types with different research output profiles can be described as follows (class sizes in per cent of the total number of projects in parentheses):

(1) Latent Class 1 '*Not Book*' (37.0%): The research output profile of this research project type is quite similar to the average profile across all projects but with fewer non-reviewed journal articles, anthologies, and monographs than the average.
(2) Latent Class 2 '*Book and Non-Reviewed Journal Article*' (35.8%): this project type uses anthologies and monographs but also non-reviewed journal articles and mass communication as primary forms

of written communication. Career development—such as diploma/degree, PhD dissertation and habilitation thesis—reviewed journal articles and follow-up projects score quite below the average.
(3) Latent Class 3 '*Multiple Outputs*' (17.9%): This project type generates research outputs in multiple ways with above-average outputs as peer-reviewed journal articles, non-reviewed journal articles, anthologies, monographs, conference papers, habilitation theses, PhD dissertations, diplomas/degrees, follow-up projects, but with fewer other conference contributions.
(4) Latent Class 4 '*Journal Article, Conference Contribution, and Career Development*' (9.3%): this most productive project type focuses strongly on peer-reviewed journal articles, with many published papers in combination with conference contributions (papers or other products), career development (diploma/degree, PhD dissertation, habilitation thesis), and follow-up projects, but this type uses fewer monographs as a form of written communication.

Of all the output variables, peer-reviewed journal articles and conference contributions discriminate the best between the LCs, with a discrimination index of about 0.60 (Table 2, last column, R^2).

5.2 Multilevel latent structure of research output profiles

In a multilevel latent structure model it is presumed that there is variation among the 22 scientific disciplines in the unconditional probabilities (the probabilities belonging to each LC). In an MLLCA the 22 scientific disciplines are grouped into latent clusters or segments according to their
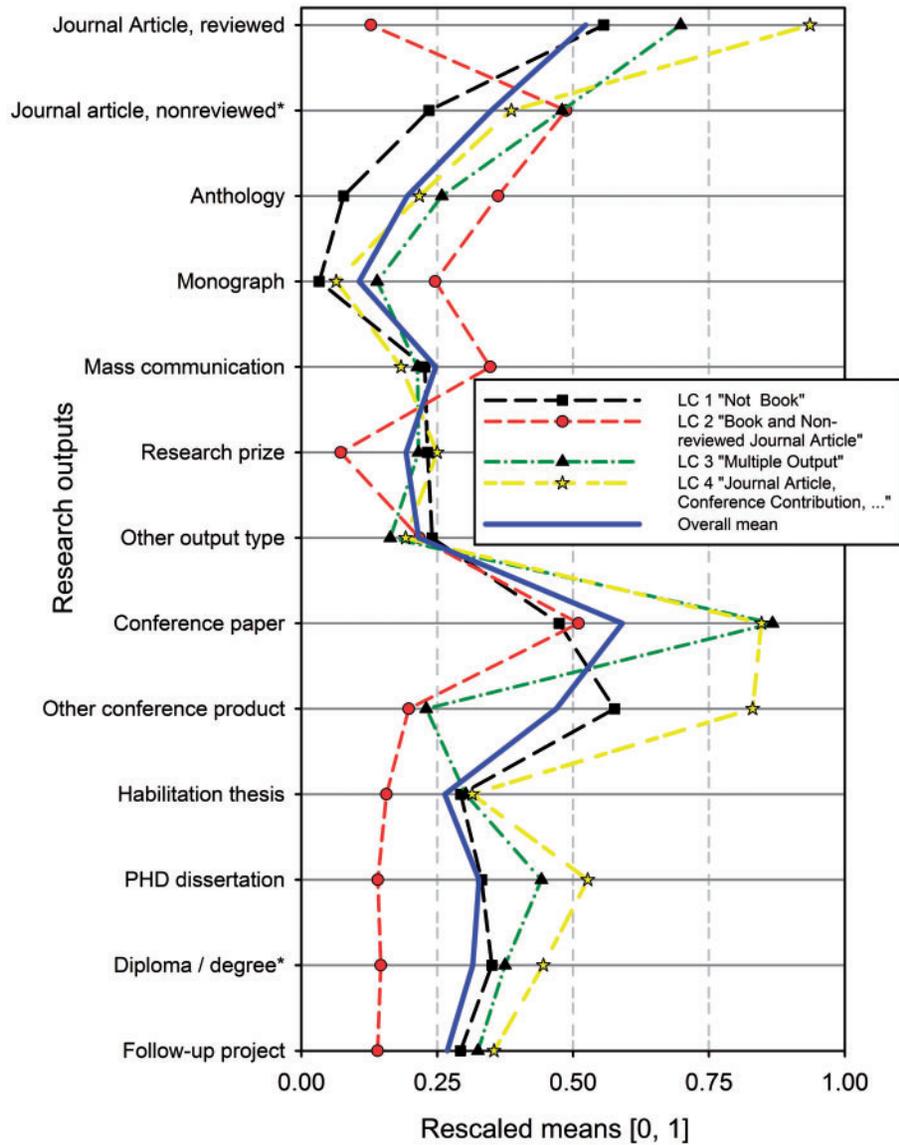
Figure 1. LCs of research output profiles (* = not used in the MLLCA).

different proportion of types of research output profiles, as obtained in Section 5.1.

Table 4 shows the results of fitting models containing one to eight latent clusters (M_1–M_8), each with four LCs and with one continuous latent C-factor, respectively. With respect to BIC(k) and AIC3, a 5-GClass model will be favoured, i.e. there are five different segments of scientific disciplines with different proportions of the project types or LCs. Additionally, using the option of 'cluster-independent C-factor', we tested (M_9) whether the same loading structure can be held in all four LCs. The BIC(k) and the AIC3 improved slightly from model M_5 to the more restricted model M_9 with $122 - 89 = 33$ fewer parameters than M_5. Therefore, the assumption of a cluster-independent C-factor held, which made it possible to compare and rank all projects on the same scale. Including direct effects, such as the association between

habilitation thesis and PhD dissertation, further improved the model. Only one residual (res = 3.88) was somewhat larger than the criterion of 3.84 (Magidson and Vermunt 2004). To fulfil the basic model assumption of local stochastic independence, we chose model M_{10} as the final model.

To assess the separation between LCs, we calculated entropy-based measures, which varied between 0 and 1.0. They show how well the observed variables were able to predict the class membership (Lukočienė, Varriale and Vermunt 2010). For LC, the $R^2_{entropy}$ amounted to 0.78, for latent clusters $R^2_{entropy}$ amounted to 0.98. The separation of both the LCs and the latent clusters is therefore very large. Another model validity index is the proportion of classification error. For each project and each LC or latent cluster a posterior probability that a project belongs to the respective class can be estimated. Out of this set of

Table 4. Fit statistics of models for variation among scientific disciplines (GClass) with four LCs and one C-factor

MNR	Models of disciplines	LL	NPAR	BIC(k)	AIC3
1	1 GClass	−17,789.4	106	35,906.4	35,896.8
2	2 GClass	−17,328.9	110	34,997.8	34,987.8
3	3 GClass	−17,211.1	114	34,774.6	34,764.2
4	4 GClass	−17,155.6	118	34,676.0	34,665.3
5	5 GClass	−17,139.7	122	34,656.4	34,645.3
6	6 GClass	−17,134.9	126	34,659.4	34,647.9
7	7 GClass	−17,133.4	130	34,668.5	34,656.7
8	8 GClass	−17,130.5	134	34,675.1	34,662.9
9	5 GClass cluster-independent C-factor	−17,188.1	89	34,651.2	34,643.1
10	Model 9 plus four additional direct effects (follow-up—PhD dissertation, habilitation thesis—PhD dissertation, habilitation thesis—anthology, monograph—anthology)	−17,166.7	93	34,620.8	34,612.4
11	Model 10 plus order restriction of the latent clusters	−17,351.5	80	34,950.2	34,943.0

Note: MNR = model number, LL = loglikelihood, NPAR = number of parameters, BIC(k) = Bayesian information criterion for k clusters, AIC3 = Akaike information criterion 3.

probabilities the highest one indicates the LC to which a project or discipline should be assigned (modal assignment). Overall, the modal assignments can deviate from the expected assignments according to the sum of the posterior probabilities. The classification error indicates the amount of misclassification. For model M_{10} the classification error was comparatively low, with 11.0% at the level of projects and 0.7% at the level of disciplines.

Based on Fig. 1 it could be supposed that the LCs do not represent a qualitative configuration but rather a quantitative dimension, in that the individual profiles run largely parallel and differ only in the level, that is, the quantity of research output. To prove this assumption the LCs were order-restricted (model M_{11}). However, the BIC(k) as well as the AIC3 of M_{11} strongly increased in comparison to all other models, with the result that the assumption of a quantitative dimension behind the LCs was not very plausible.

To illustrate the meaning of these segments of scientific disciplines, Table 5 shows the distribution of the projects among the four LCs (Fig. 1) of each of the five segments of disciplines (latent clusters). The last column of numbers in Table 5 indicates the size of the LCs or types of research

output profiles. The last row of numbers in Table 5 indicates the proportion of disciplines that were in each discipline segment. The latent clusters or segments of scientific disciplines can be described according to the disciplines that belong to them (cluster sizes in per cent of the total number of disciplines in parentheses):

- Latent Cluster 1 '*Life Sciences and Medicine*' (31.6%): biology; botany; zoology; geosciences; preclinical medicine; clinical medicine; agricultural, forestry and veterinary sciences.
- Latent Cluster 2 '*Social Sciences/Arts and Humanities*' (31.4%): social sciences; jurisprudence; philosophy/theology; history; linguistics and literary studies; art history; other humanities fields.
- Latent Cluster 3 '*Formal Sciences*' (13.9%): mathematics; computer sciences; economic sciences.
- Latent Cluster 4 '*Technical Sciences*' (13.5%): Other natural sciences; technical sciences; psychology.
- Latent Cluster 5 '*Physical Sciences*' (9.6%): physics, astronomy and mechanics; chemistry.

The remaining columns in Table 5 show the distribution of projects in each discipline segment or the probability of a project showing a specific profile type given its latent cluster membership. For instance, of all projects falling into the first GClass 84% are in LC 1 ('Not Book'), 0% are in LC 2 ('Book and Non-Reviewed Journal Article'), 6% are in LC 3 ('Multiple Outputs'), and 10% are in LC 4 ('Journal Article, Conference Contribution, and Career Development'). High proportions in a cell indicate a strong association of the corresponding segment of disciplines in the column with the corresponding type of research output profile in the row. In this respect the segment 'Life Sciences and Medicine' (GClass 1) was strongly associated with the 'Not Book' project type (LC 1) (84% of projects of this segment), but 10% of this cluster fell also in the most productive type, 'Journal Article, Conference Contribution, and Career Development' (LC 4). In the segment 'Social Sciences/Arts and Humanities' (GClass 2) almost all projects (97%) are of the second 'Book and Non-Reviewed Journal Article' type (LC 2). Projects of the third segment 'Formal Sciences' are classified about 80% in the 'Multiple Outputs' type, 14% also in the 'Not Book' type. The fourth segment, 'Technical Sciences', is rather heterogeneous, with over 95% of the projects of this segment in the first three project types and 37% even in the 'Book and Non-Reviewed Journal Article' type (LC 2). The projects of the last segment, 'Physical Sciences', can be divided mainly into two groups: 38% in the first project type 'Not Book' and 56% in the most productive project type, 'Journal Article, Conference Contribution, and Career Development'. Overall, except for 'Humanities', there is no one-to-one assignment of a segment of disciplines to a special type of research output profile.

Table 5. Relative class sizes and distribution of projects among LCs (project output types) within each latent clusters (discipline segments) for M_{10} (column per cent)

Latent classes (research output profile types)	Latent clusters (discipline segments)					LC size
	GClass 1	GClass 2	GClass 3	GClass 4	GClass 5	
LC 1 'Not Book'	**0.84**	0.00	0.14	**0.35**	**0.38**	0.37
LC 2 'Book and Non-Reviewed Journal Article'	0.00	**0.97**	0.02	**0.37**	0.00	0.36
LC 3 'Multiple Outputs'	0.06	0.03	**0.81**	0.24	0.06	0.18
LC 4 'Journal Article, Conference Contribution, Career Development'	0.10	0.00	0.03	0.04	**0.56**	0.09
GClass size	0.32	0.31	0.14	0.14	0.10	

Note: LC size = size of the latent class, GClass size = size of the latent clusters, proportions over 0.30 (except for class sizes) are in **bold face**.

Disciplines show great heterogeneity in their research output profiles.

Figure 2 shows the LC proportions for each single discipline, structured according to the latent cluster (segments of disciplines). This finding also replicated the basic findings in Table 5 at the level of single disciplines. It is of interest that the 'Book and Non-reviewed Journal Article' type (LC 2) played an important role not only in 'Social Sciences/Arts and Humanities' but also in 'Technical Sciences'.

5.3 Explaining LC membership

To explain the LC membership we conducted a modified multilevel multinomial regression model with the latent-class membership as categorical variable and the set of covariates as predictors (Vermunt 2010). Beforehand, the continuous covariates time, age, duration, overall rating of a proposal (ex ante evaluation), and requested grant sum were z-transformed ($M = 0$, $S = 1$) to facilitate the interpretation of the regression results independently of the units of the covariates (Table 6).

Wald statistics are used to assess the statistical significance of a set of parameter estimates. Using Wald statistics, the restriction is tested that each estimate in a set of parameters associated with a given covariate equals zero (Vermunt and Magidson 2005b). A non-significant Wald statistic indicates that the respective covariate does not differ between the LCs. Additionally, we calculated a z-test for each single parameter. There are three covariates that explained the class membership with statistically significant Wald tests: project duration, requested grant sum, and the project head's age. The overall rating of the proposal (ex ante evaluation), for instance, had no impact on the class membership. Research projects with a duration longer than the average of 39 months were more often in LC 4 ('Journal Article, Conference Contribution, and Career Development') than research projects with a shorter than average duration were. The higher the requested grant sum of a project, the less probable it was for the project to be in LC 2 ('Book and

Non-Reviewed Journal Article'), but the more probable it was for it to be in LC 4 ('Journal Article, Conference Contribution, and Career Development'). Projects where the project head was older than the average age of 47 were more frequently in LC 2 ('Book and Non-Reviewed Journal Article'), whereas projects where the project head was younger than 47 tended to be in LC 3 ('Multiple Outputs'). Additionally, the percentage of projects in LC 4 ('Journal Article, Conference Contribution, and Career Development') decreased from project end year 2002 to project end 2010.

In sum, projects that belong to the 'Book and Non-Reviewed Journal Article' type (LC 2) tended to have rather low requested grant sums and project heads who were older than the average, whereas the most productive 'Journal Article, Conference Contribution, and Career Development' type was characterized by above-average requested grant sums and above-average project durations. Further, the percentage of this most productive type decreased over time (time of project end). The third type, 'Multiple Outputs', tended to have younger project heads.

5.4 Ranking of projects

Until now it was assumed that output profiles of research projects can be fully explained by the LC or types of output profiles into which the projects were classified. However, as Table 3 shows, projects differed not only with respect to LCs or latent cluster but also with respect to an additional quantitative dimension, a latent C-factor, referring to classical concepts of factor analysis. Unlike LCs, all output variables have positive loadings on this dimension—namely, with the same correlation or loading structure within each LC. Thus, the higher the value in any of the output variable, the higher the value of the C-factor is. Positive values in the C-factor represent productivity above average of the projects in this LC, and negative values indicate projects with less productivity with respect to projects in the same LC. In sum, the C-factor represents productivity differences of projects within each
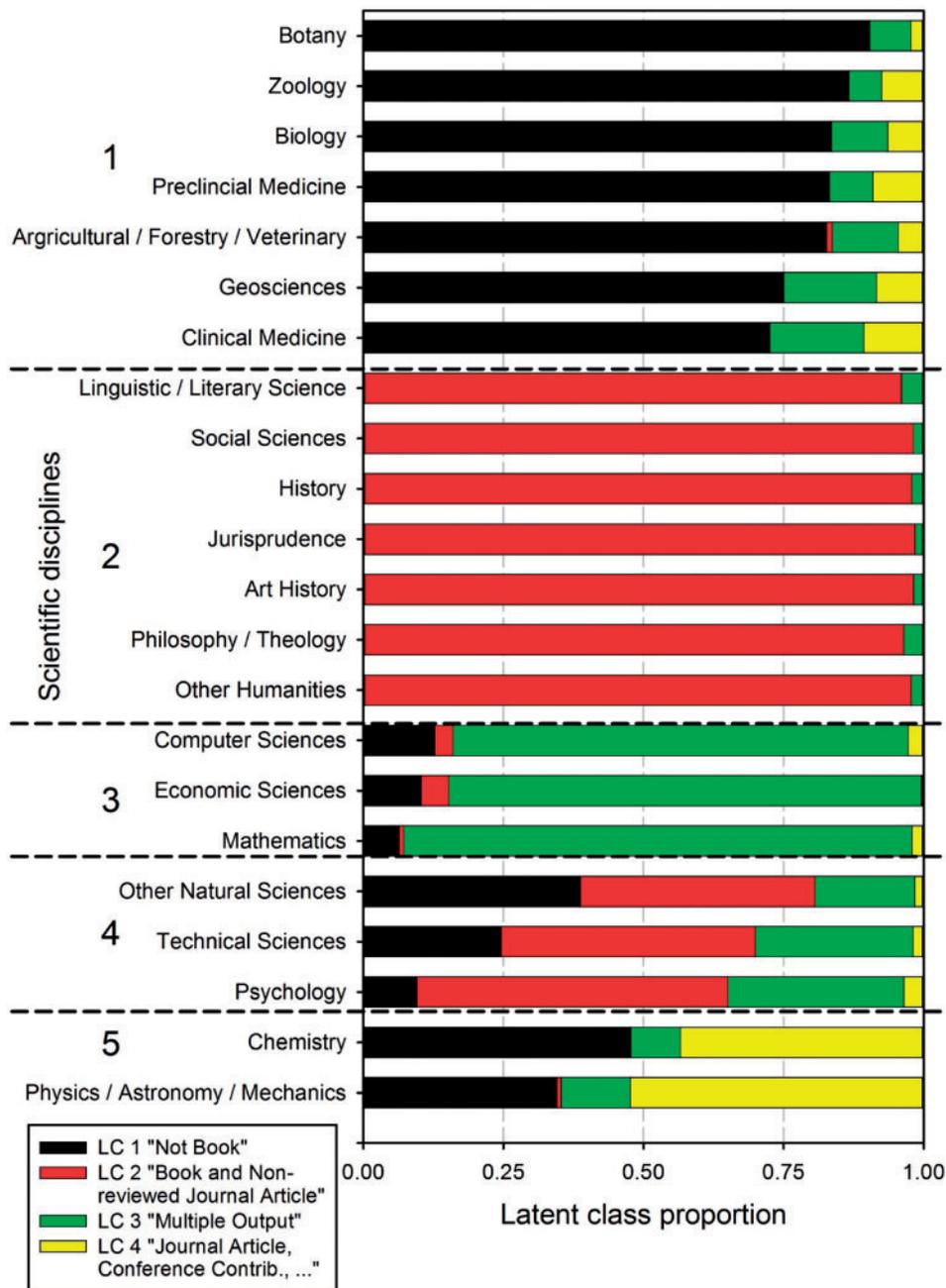
Figure 2. Estimated proportions of the four LCs of projects for each scientific discipline (stacked bars plot), classified into one of five latent clusters (1–5, separated by dashed lines).

LC, similar to a Mixed-Rasch model in psychometrics (Mutz, Borchers and Becker 2002; Mutz and Daniel 2007). This type of ranking can be used by the FWF (and other funding organizations) for comparative evaluation of the output of different projects within a certain time period.

According to the C-factor, the projects within each LC or project type could be ranked (Fig. 3) from left (projects with the highest productivity) to right (projects with the lowest productivity). Additionally, Goldstein-adjusted confidence intervals are shown which makes it possible to interpret non-overlapping intervals of two projects as statistical significant differences at the 5% probability level (Mutz and Daniel 2007). Roughly speaking, only the first and the last 100 projects in each LC actually showed statistically significant differences in their C-factor values.

6. Discussion

The aim of this study was to conduct a secondary analysis of final report data from the FWF (ex post evaluation) for
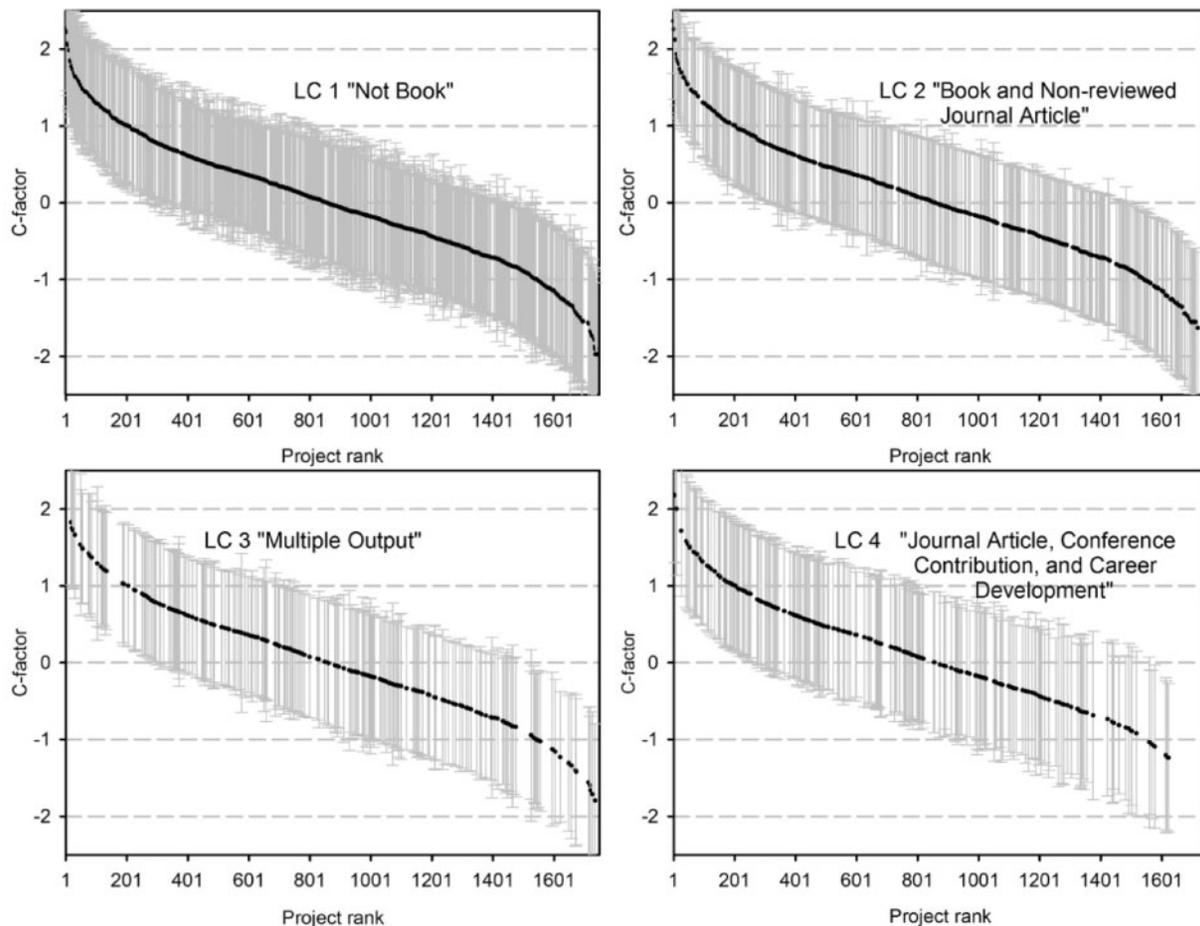
Figure 3. Rankings of projects within LCs from left (largest amount of research output) to right (smallest amount of research output) and Goldstein-adjusted confidence intervals.

the years 2002–10 (project end) and—using multilevel LCA—to build bottom-up a typology of research projects and, further, to classify scientific disciplines according to the different proportions of the types of research output profiles found. Referring to our four research questions, the results can be summarized as follows:

(1) The 1,742 completed FWF-funded research projects available for a final report can be classified according to the research output profiles in the following four types with relatively high discrimination: 37% of all projects are in the 'Not Book' type, 35.8% in the 'Book and Non-Reviewed Journal' type, 17.9% in the 'Multiple Outputs' type, and 9.3% in the 'Journal Article, Conference Contribution, and Career Development' type, which is the most productive type in terms of number of journal articles and career-related activities. These project types represent primarily a qualitative configuration and not a quantitative dimension according to which projects can be ranked.

(2) The 22 scientific disciplines can be divided into five segments of disciplines based on different proportions of the types of research output profiles: 31.6% of all projects can be classified in the segment 'Life Science and Medicine', 31.4% in 'Social Sciences/Arts and Humanities', 13.9% in 'Formal Sciences', 13.5% in 'Technical Sciences' and 9.6% in 'Physical Sciences', such as chemistry and physics. Only the 'Social Sciences/Arts and Humanities' segment is almost fully associated with one research output profile ('Book and Non-Reviewed Journal Article' type); all other segments show different proportions of the four research output profiles. Psychology and economic sciences are usually subsumed under humanities and social sciences. But the MLLCA showed that these two scientific disciplines do not belong to the segment 'Social Sciences/Arts and Humanities'. Additionally, the fourth and most productive type of research output profile is highly represented (56%) in the fifth segment of disciplines, 'Physical Sciences', and with only 10% in 'Life Science and Medicine', contrary to the findings of the DFG (Deutsche

Table 6. Selected model parameters of the regression from LCs on covariates

Covariate	Latent classes								Overall test Wald
	LC 1		LC 2		LC 3		LC 4		
	Not Book		Book and Non-Reviewed Journal Article		Multiple Outputs		Journal Article, Conference Contribution, Career Development		
	Par	SE	Par	SE	Par	SE	Par	SE	
Time period of the approval decision	−0.11	0.63	−0.85	1.09	−1.05	0.91	2.01	1.05	3.73
Time period of the project end	0.40	0.63	1.01	1.05	0.81	0.90	−2.22*	1.09	4.26
Project duration	−0.11	0.29	−0.93	0.47	−0.52	0.40	1.56*	0.51	9.62**
Overall rating of the proposal	−0.19	0.15	−0.16	0.25	−0.04	0.21	0.40	0.26	3.53
Requested grant sum	−0.28	0.19	−1.17*	0.37	0.45	0.26	1.00*	0.28	23.90**
Project head's sex	0.51	0.61	−0.10	0.97	−0.72	1.12	0.30	0.76	0.77
Project head's age	−0.25	0.13	0.72*	0.23	−0.49*	0.22	0.02	0.21	13.59**

Note: LC = latent class, Par = parameter estimate, SE = standard error, Wald = Wald test, df = degrees of freedom.
*$p < 0.05$ (z-test) **$p < 0.05$ (Wald test, df = 3).

Forschungsgemeinschaft 2005) mentioned above in the introduction. 'Life Sciences and Medicine' is strongly associated (84%) with the 'Not Book' type. Projects of the third segment, 'Formal Sciences', are classified about 80% in the 'Multiple Outputs' type and 14% also in the 'Not Book' type. The fourth segment, 'Technical Sciences', is rather heterogeneous, with over 90% of the projects in this segment in the first three project types and 37% even in the 'Book and Non-Reviewed Journal Article' type.

In the end, the findings of the Expert Group on Assessment of University-Based Research set up by the European Commission (European Commission 2010) on the disciplines' preferred forms of communication are too simple. To sum up, there are not only differences between scientific disciplines in the research output profiles; there is also great heterogeneity of research output profiles within disciplines and segments of disciplines, respectively.

(3) Membership in a particular project type can essentially be explained by three covariates—project duration, requested grant sum, and the project head's age. Projects that belong to the 'Book and Non-Reviewed Journal Article' type tend to be characterized by small requested grant sums and project heads who are older than the average, whereas the most productive type, 'Journal Article, Conference Contribution, and Career Development', tends to be characterized by high requested grant sums and longer than average project durations, but whose proportion decreases the more the date of the project termination approximates the year 2010. Reviewers' overall rating of the proposal (ex ante evaluation) had no influence on latent-class membership.

(4) Projects differ not only in the qualitative configuration of research outputs, their research output profiles, but also with respect to a quantitative dimension that makes productivity rankings of projects possible. The higher the output of a project in each of the research output variables, the higher its value on the quantitative (latent) dimension is. Only the first and the last 100 projects within each project type differed statistically significantly on this dimension.

However, there are also some limitations of our study that have to be discussed: first, the findings represent a specific picture of the research situation in one country, namely, Austria, in a 10-year period situation, and they may not necessarily apply in other countries. The quality of the research was not considered, such as through using international reference values for bibliographic indicators (Opthof and Leydesdorff 2010; Bornmann and Mutz 2011) or through using discipline-specific quality criteria. Second, the study included only projects (in particular, 'Stand-Alone Projects') that were funded by the FWF. Research projects in Austria that were funded by other research funding organizations, that were not Stand-Alone Projects (40%) or that were funded by higher education institutions themselves could not be included. Further, research projects are mostly financed by mixed funding—that is, in part by grants from various research funding organizations and in part by matching funds from the relevant higher education institution (e.g. human resources), so that research output profiles cannot necessarily be explained by covariates of a single research funding organization. Third, the persons responsible for preparing a report (here, the project heads) always have a certain leeway to mention or not mention certain results of their

research as results of the FWF-funded research projects in the final report (e.g. journal articles, career development). In social psychology terms, this phenomenon can be subsumed under the concept of 'social desirability' (Nederhof 1985). Social desirability is a psychological tendency to respond in a manner that conforms to consensual standards and general expectancies in a culture. The findings of this study could thus also in part reflect different report policies in the different scientific disciplines.

7. Conclusions

Despite these limitations, we draw the following conclusions from the results:

(1) *Concept of 'research output':* If the aim is to include all disciplines in the ex post research evaluation, it is necessary to define the term 'research output' more broadly, as do the RCUK and the FWF, and to include—in addition to journal articles—also other output categories, such as monographs, anthologies, conference contributions, and patents, in order to treat all disciplines fairly with regard to research output.

(2) *Arts and Humanities*: As has been repeatedly demanded, the arts and humanities really should be treated as an independent and relatively uniform area (Nederhof et al. 1989; Nederhof 2006). Instead of counting only journal articles and their citations, however, it is important to include also monographs and anthologies (Kousha and Thelwell 2009). Psychology and economic sciences do not belong to the segment 'Social Sciences/Arts and Humanities'. Therefore, it is rather problematic to subsume psychology, economic sciences, social sciences, sociology, and humanities in one unique concept, 'Social Sciences and Humanities', as is often the case (Archambault et al. 2006; Nederhof 2006).

(3) *Hierarchy of the sciences*: A most familiar and widespread belief is that scientific disciplines can be classified as 'hard' sciences and 'soft' sciences, with physics at the top of the hierarchy, social sciences at the bottom and biology somewhere in between (Smith et al. 2000). The strategy followed here made it possible to work out, bottom-up from the research outputs of funded research projects, an empirically based typology of scientific disciplines that at its heart is not hierarchically structured. The typology found reflects much more strongly the real structure of science than the top-down classification systems of sciences allow. However, the identified research output profiles do not unambiguously indicate the segment of the discipline. For instance, almost all projects in the segment 'Social Sciences/Arts and Humanities' are of the 'Book and Non-Reviewed Journal Article' type, but not all projects of the 'Book and Non-Reviewed Journal Article' type are in the segment 'Social sciences/Arts and Humanities'; there is also a high proportion of 'Book and Non-Reviewed Journal Article' type projects in the segment 'Technical Sciences'.

(4) *Research output profiles:* Using MLLCA, research projects are not examined with regard to few arbitrarily selected project outputs; instead, the profile or combination of multiple research outputs is analysed. This should receive more attention also in ex post research evaluations of projects.

(5) *Ranking of projects*: In addition, with MLLCA a qualitative dimension of different types of projects and segments of disciplines can be distinguished from a quantitative dimension that captures research productivity. In this way, projects and possibly also scientific disciplines can be ranked according to their productivity.

References

Archambault, E. *et al.* (2006) 'Benchmarking Scientific Output in the Social Sciences and Humanities: The Limits of Existing Databases', *Scientometrics*, 68/3: 329–42.

Banzi, R. *et al.* (2011) 'Conceptual Frameworks and Empirical Approaches Used to Assess the Impact of Health Research: An Overview of Reviews', *Health Research Policy and Systems*, 9: 26.

Bence, V. and Oppenheim, C. (2005) 'The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions', *Journal of Educational Administration and History*, 35/2: 137–55.

Bijmolt, H. A., Pass, L. J. and Vermunt, J. K. (2004) 'Country and Consumer Segmentation: Multi-Level Latent Class Analysis of Financial Product Ownership', *International Journal of Research in Marketing*, 21: 323–40.

Bolck, A., Croon, M. A. and Hagenaars, J. A. (2004) 'Estimating Latent Structure Models with Categorical Variables: One-Step Versus Three-Step Estimators', *Political Analysis*, 12: 3–27.

Bornmann, L. (2012) 'Measuring the societal impact of research', *EMBO Reports*, 13/8: 673–6.

Bornmann, L. and Mutz, R. (2011) 'Futher Steps Towards an Ideal Method of Measuring Citation Performance: The Avoidance of Citation (ratio) Averages in Field-Normalization', *Journal of Informetrics*, 5/1: 228–30.

Bourke, P. and Butler, L. (1996) 'Publication Types, Citations Rates and Evaluation', *Scientometrics*, 37/3: 473–94.

Bozdagon, H. (1993) 'Choosing the Number of Component Clusters in the Mixture-Model Using a New Informational Complexity Criterion of the Inverse-Fisher Information Matrix'. In: Opitz, O., Lausen, B. and Klar, R. (eds) *Information and Classification*, pp. 218–34. Germany: Springer, Heidelberg.

Butler, L. (2008) 'ICT Assessment: New Measures of Biomedical Research Impact', *Scientometrics*, 74/1: 39–55.

Butler, L. and McAllister, I. (2009) 'Evaluating University Research Performance Using Metrics', *European Political Science*, 10: 44–58.

——. (2011) 'Metrics or Peer Review? Evaluating the 2001 UK Research Assessment Exercise in Political Science', *European Political Review*, 7: 3–17.

Buxton, M. and Haney, S. (1998) 'Evaluating the NHS Research and Development Programme: Will the Programme Give Value for Money?', *Journal of the Royal Society of Medicine*, 35/91: 2–6.

Cronin, B. and La Barre, K. (2004) 'Mickey Mouse and Milton: Book Publishing in the Humanities', *Learned Publishing*, 17: 85–98.

Deutsche Forschungsgemeinschaft. (2005) *Publishing Strategies in Transformation? Results of a Study on Publishing Habits and Information Acquisition with Regard to Open Access.* Weinheim, Germany: Deutsche Forschungsgemeinschaft.

Dinges, M. (2005) *The Austrian Science Fund: Ex Post Evaluation and Performance of FWF Funded Research Projects.* Vienna: Institute of Technology and Regional Policy.

European Commission: Expert Group on Assessment of University-Based Research. (2010) *Assessing Europe's University Based Research*, European Commission Directorate-General for Research <http://www.eua.be/eua-work-and-policy-area/research-and-innovation/European-Commission-Expert-Groups.aspx > accessed 3 Nov 2012.

European Science Foundation. (2011a) *The Capture and Analysis of Research Outputs*, <http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html > accessed 3 Oct 2012.

——. (2011b) *The Classification of Research Portfolios*, <http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html> accessed 3 Oct 2012.

Fischer, C. and Reckling, F. (2010) *Factors Influencing Approval Probability in FWF Decision-Making Procedures.* Vienna, Austria: Austrian Science Fund (FWF).

Glänzel, W. (1996) 'A Bibliometric Approach to Social Sciences: National Research Performance in 6 Selected Social Science Areas 1990-1992', *Scientometrics*, 35: 291–307.

Glass, G. V. (1976) 'Primary, Secondary, and Meta-Analysis of Research', *Educational Researcher*, 5: 3–8.

Gonzalez-Brambila, C. and Velosos, F. M. (2007) 'The Determinants of Research Output and Impact: A Study of Mexican Researchers', *Research Policy*, 36: 1035–51.

Hicks, D. (2004) 'The Four Literatures of Social Sciences'. In: Moed, H. (ed.) *Handbook of Quantitative Science and Technology Research*, pp. 476–96. Dordrecht, The Netherlands: Kluwer.

Huang, M. and Chang, Y. (2008) 'Characteristics of Research Output in Social Sciences and Humanities: From a Research Evaluation Perspective', *Journal of the American Society for Information Science and Technology*, 59/11: 1819–28.

Kimberly, K. L. and Muthén, B. (2010) 'Multilevel Latent Class Analysis: An Application of Adolescent Smoking Typologies with Individual and Contextual Predictors', *Structural Equation Modeling*, 17: 193–215.

Kousha, K. and Thelwell, M. (2009) 'Google Book Search: Citation Analysis for Social Science and the Humanities', *Journal of the American Society for Information Science and Technology*, 60/8: 1537–49.

Lewison, G. (2003) 'Beyond Outputs: New Measures of Biomedical Research Impact', *Aslib Proceedings*, 55/1-2: 32–42.

Linmans, A. J. M. (2010) 'Why with Bibliometrics the Humanities does not Need to be the Weakest Link - Indicators for Research Evaluation based on Citations, Library Holdings, and Productivity Measures', *Scientometrics*, 83: 337–54.

Lukočienė, O. and Vermunt, J. K. (2010) 'Determining the Number of Components in Mixture Models for Hierarchical Data'. In: Fink, A. *et al.* (eds) *Advances in Data Analysis, Data Handling and Business Intelligence*, pp. 241–9. Heidelberg, Germany: Springer.

Lukočienė, O., Varriale, R. and Vermunt, J. K. (2010) 'The Simultaneous Decision(s) about the Number of Lower- and Higher-Level Classes in Multilevel Latent Class Analysis', *Sociological Methodology*, 40/1: 247–83.

Magidson, J. and Vermunt, J. K. (2004) 'Latent Class Models'. In: Kaplan, D. (ed.) *The Sage Handbook of Quantitative Methodology for the Social Sciences*, pp. 175–98. Thousand Oaks, CA, USA: Sage.

McCutcheon, A. L. (1987) *Latent Class Analysis.* London, UK: Sage.

Mutz, R. and Daniel, H.-D. (2007) 'Entwicklung eines Hochschul-Rankingverfahrens mittels Mixed-Rasch-Modell und Mehrebenenanalyse am Beispiel der Psychologie [Development of a Ranking Procedure by Mixed Rasch Model and Multilevel Analysis - Psychology as an Example]', *Diagnostica*, 53/1: 3–16.

——. (2012) 'University and Student Segmentation: Multilevel Latent-Class Analysis of Students' Attitudes Towards Research Methods and Statistics', *British Journal of Educational Psychology*, DOI: 10.1111/j.2044-8279.2011.02062.x.

Mutz, R. and Seeling, U. (2010) 'A nomothetic Version of the Brunswikean Lens Model: A Variable- and Person-Oriented Approach', *Zeitschrift für Psychologie/Journal of Psychology*, 218/3: 175–84.

Mutz, R., Borchers, J. and Becker, G. (2002) 'Analysis of current and potential commitment of private forest owners in North Rhine-Westphalia based on the mixed Rasch model', *Forstwissenschaftliches Centralblatt*, 121/1: 35–48.

Mutz, R., Bornmann, L. and Daniel, H.-D. (2012a) 'Does Gender Matter in Grant Peer Review? An Empirical Investigation Using the Example of the Austrian Science Fund', *Journal of Psychology*, 220/2: 121–9.

——. (2012b) 'Heterogeneity of inter-rater reliabilities of grant peer reviews and its determinants: a general estimating equations approach', *PLoS ONE*, 7/10: e48509.

Nederhof, A. J. (1985) 'Methods of Coping with Social Desirability Bias - A Review', *European Journal of Social Psychology*, 15/3: 263–80.

——. (2006) 'Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66/1: 81–100.

Nederhof, A. J. *et al.* (1989) 'Assessing the Usefulness of Bibliometric Indicators for the Humanities and the Social and Behavioural Sciences: A Comparative Study', *Scientometrics*, 15/5-6: 423–35.

Opthof, T. and Leydesdorff, L. (2010) 'Caveats for the Journal and Field Normalizations in the CWTS ("Leiden") Evaluations of Research Performance', *Journal of Informetrics*, 4/3: 423–30.

Rindskopf, D. (2006) 'Heavy Alcohol Use in the "Fighting Back" Survey Sample: Separating Individuals and Community Level Influences Using Multilevel Latent Class Analysis', *Journal of Drug Issues*, 6/2: 441–62.

Sarli, C., Dubinsky, E. K. and Holmes, K. L. (2010) 'Beyond Citation Analysis: A Model for Assessment of Research Impact', *Journal of the Medical Library Association*, 98/1: 17–23.

Smith, L. D. *et al.* (2000) 'Scientific Graphs and the Hierarchy of Sciences: A Latourian Survey of Inscription Practices', *Social Studies of Science*, 30/1: 73–94.

Sturn, D. and Novak, R. (2012), Was können Forschungsevaluationen durch den FWF zur Entwicklung von Hochschulen beitragen?' [What Can FWF's Research Evaluations Contribute to the Development of Institutions of Higher Education?]. In: Österreichische Qualitätssicherungsagentur

(ed.) Braucht Forschung Qualitätsmanagement?, pp. 59–70. Vienna, Austria: AQA.

Van Leeuwen, T. (2006) 'The Application of Bibliometric Analyses in the Evaluation of Social Science Research. Who Benefits from it, and Why it is Still Feasible', *Scientometrics*, 66/1: 133–54.

Vermunt, J. K. (2003) 'Multilevel Latent Class Models', *Sociological Methodology*, 33/1: 213–39.

——. (2010) 'Latent Class Modeling with Co-Variates: Two Improved Three-Step Approaches', *Political Analysis*, 18: 450–69.

Vermunt, J. K. and Magidson, J. (2005a) *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont, MA, USA: Statistical Innovations Inc.

——. (2005b) Latent GOLD 4.0 User's Guide, <http://www.statisticalinnovations.com/technicalsupport/LGusersguide.pdf> accessed 3 Oct 2012.

White, H. D. *et al.* (2009) 'Libcitations: A Measure of Comparative Assessment of Book Publications in the Humanities and Social Sciences', *Journal of the American Society for Information Science and Technology*, 60/6: 1083–96.