Specialized Research Output and Performance

Harris Dellas,[a,c,d] Manu García[b,a] Carlos Garriga[a] Christian Zimmermann[a]
[a]Federal Reserve Bank of St. Louis
[b]Washington University in St. Louis
[c]Karl Brunner Institute
[d] University of Surrey

Hydra 24, Sept 2024

The questions

A set of –novel– questions pertaining to the implications of the **choice of the composition of research portfolio**

- Does the degree of specialization of research output matter for performance (level/quality of output)?
- Does its time profile (early vs late in the career) matter?
- If output variety is rewarded, are there any simple ways to pursue it?

The obstacles

(At least) a couple of major challenges stand in the way

- What is a useful –measurable and consequential– definition of the degree of specialization?
  Virgin territory in the case of research activities
- How to identify its causal effects on performance?
  The difficulty here lies in the fact that one's choice of research diversification may be influenced by things (such as ability and effort) that matter for performance on their own. This possibility makes it hard to isolate the contribution of specialization.

The measurement challenge

- Trade theory provides a framework for thinking about a production specialization index
  Economic activities are classified in sectors (at various levels of aggregation). Their shares in total economic activity -usually in comparison to other countries- are used to construct a measure of a country's degree of specialization
- In an analogous fashion, research works can be classified in areas/topics and the frequencies of the areas appearing in one's portfolio of works can be used to construct a measure of specialization
- This is feasible because, uniquely among all sciences, economics has such classification systems: the JEL and the NEP schemes
- A portfolio of research works is simply a collection of JEL codes and its specialization index is a function of those JELs

Measurement: A specialization index

- Our main index of specialization is an HH index, with the portfolio being the market and the individual JELs or NEPs being the firms.
- Another is the count of distinct JELs in a portfolio
- As JEL codes come in categories and subcategories , it is possible to construct local (narrow) and broad specialization indexes

Assessing the causal effect of specialization on performance

- Estimate the effect of the specialization index after accounting for (conditioning on) other determinants of performance
- The difficulty: The main determinants of performance (ability-human capital and effort) are unobservable and may also matter for the specialization choice
- We use various proxies of ability-effort (academic pedigree, Num. of top 5, top publishing authors within individual academic cohorts etc.) to overcome this difficulty

Main findings

- Q1. Greater diversification leads to better performance (quality adjusted pages, quality adjusted citations, citations from work in other fields , etc.) and higher rank. Significant in terms of both t-stat and R^2 and true for all of groups of researchers considered
- Q2. Greater early (5-7 years from the time of the first publication) diversification lowers the probability of research attrition
  Tentative: But for a given level of lifetime diversification, those who specialized more early on, seem to fare worse
- Q3. Co-authorship is a common way to achieve greater output diversification

A bit of theory

- Trade theory: Under CRS or IRS, complete specialization maximizes the value of output.
- It represents optimal strategy under risk neutrality or certainty or complete asset markets.
- With incomplete asset markets and risk aversion, insurance considerations favor -some- diversification at the cost of lower production-consumption
- But are research activities similar to other economic activities?
- Research production has features that are missing from the activities considered elsewhere (e.g. international trade) and which may favor diversification

How research and goods production differ?

- **Joint work**: Production in trade models does not take into account the role of joint work: Co-authorship, due to complementary skills/ expertise, can support more diversified- higher quality output (higher TFP)
- **Cross-fertilization**: Undertaking research in diverse areas exposes one to different practices and problems, facilitating the transfer of knowledge (methods, perspective, –) and potentially increasing productivity
- An example: In the Manhattan project, Oppenheimer instituted a weekly colloquium for hundreds of scientists. He understood the value of gathering people from different parts of a project in the same place, encouraging them to discuss their work and combine their ideas. As Bethe commented, "*..very often a problem discussed in one of these meetings would intrigue a scientist in a completely different branch of the laboratory, and he would come up with unexpected solutions*."

The RePEc data

- Data regarding who authored which paper (working paper/journal article/book chapter), where it was published, who reads it and where it is cited
- The -bibliographic- data (called metadata) come from publishers (commercial or academic) using a common format. They are provided through their servers, which anybody can access. RePEc is simply a scheme to organize metadata and make them available in the public domain (RePEC was founded in 1997).
- Only journals and working paper series that are listed in RePEc are classified, and only authors that have registered themselves are included.

RePEc ranking

Compared to other ranking exercises:

1. RePEc is very comprehensive
2. includes some criteria that are unique (such as readership, # of authors citing, centrality among co-authors,...)
3. includes working papers
4. is internally consistent-self contained –the same source is used both to establish impact factors of publication venue and rankings of authors or institutions. Ranking exercises for institutions/authors rely heavily on journal-publication venue **impact factors** calculated elsewhere. These impact factors are usually the most critical and controversial issue with rankings. RePEc impact factors are determined with the RePEc data. They are determined/evolve in real time

RePEc: Impact factors/journal rankings

Several methods of calculating impact factors

- Simple Impact Factor (eg. # total citations to AER divided by # items in AER)
- Recursive Impact Factor: similar to the Google Page-Rank which ranks web pages higher if they are linked to many others, even more so if the web sites they are linked to have a high Page-Rank themselves; RePRc does this for every journal and paper series instead of each web page)
- Discounted Impact Factor (what is hot now)
- Recursive Discounted Impact Factor
- Journal H-Index
- Abstract Views, Downloads

Aggregation: Harmonic mean of the various impact factors

Author rankings

Uses multiple criteria from various categories

- Number of Works
- Citation Counts
- Journal Page Counts
- Popularity on RePEc (downloads, abstract views)
- Co-Authorship Networks (pre-eminence)

Author ranking

Aggregation: **Harmonic Mean of Author Rank, M, in each of 33 author ranking criteria** $M = \frac{N}{\sum_{i=1}^{N} \frac{1}{r_i}}$, where r_i author ranking in criterion i.

- Very good rankings have more weight; for example, the first rank counts twice as much as the second one
- But rank difference carries little weight for higher numbers (lower ranked)
- This aggregation rewards those who are particularly good in some category
- (perhaps a bit too much: to dampen this, the RePEc ranking adds a constant to each rank and then subtracts it from the mean)

Also, rankings by field can be constructed on the basis of NEPs

Classification schemes

- The **JEL system** has 26 primary categories (spanning the letters of the English alphabet), each one with secondary and tertiary subcategories. JEL codes A through Z denote primary categories (e.g. **F** International Economics); secondary categories are specified by an additional Arabic numeral (e.g. **F1** International Trade), and tertiary categories by a following second Arabic numeral (e.g. **F18** Trade and Environment). Many published papers in economics carry one or more JEL code, provided by the authors (but often corrected by the editors of the journal)
- The **NEP system** has 100 field classifications

A diversification-specialization measure

- We propose a specialization-diversification index (SDI) inspired by the Herfindahl-Hirschman Index (HHI), a common measure of market concentration
- The researcher's portfolio is the market, and the individual JEL codes are the firms
- Instead of the distribution of the market share of firms, we have the distribution of the JEL codes in the portfolio
- $SDI = \frac{\sum_j i^2}{(\sum_j i)^2}$, where i is the number of papers that contain code j
- A lower value of SDI represents a greater degree of diversification: The larger the number of JEL fields in the portfolio of works and the more even their distribution, the lower the SDI score

Empirical Analysis

The data

- RePEc contains over 4.5 million research items from 4,000 journals and 5,500 working paper series, with over 65,000 authors registered
- Cross section data set: Lifetime performance measures and ranking of registered authors in the latest RePEc release (April 2024)
- Annual, unbalanced panel with performance measures from 2005 up to the latest release

Variables

Diversification indexes

1. *SDI* uses full (letter plus both digits) JEL code
2. $SDI - L1$ uses letter plus first digit of the JEL code
3. $SDI - L$ uses only the letter part of the JEL code
4. *NEP* uses NEP classification system

Broad vs narrow diversification measures. Interdisciplinary vs local diversification

The data

Variables

- Rank or score, R
- Year of graduation, as recorded in the RePEc Genealogy
- Graduating institution's *current* rank (author pedigree)
- Year of first published item recorded in RePEc (academic age)
- Works weighted by recursive impact factor, WScWorks
- Citations weighted by recursive impact factor of citing publication, WScCites
- Journal article pages weighted by recursive impact factor of journal, WScPages
- Number of NEP fields citing author, NEPCites
- # of items (total, with JEL code, with NEP code)
- Gender
- Total number of co-authors (RePEc registered and unregistered)

Results: Diversification and rankings

Table 1: Diversification and performance: Full set of JEL codes

	Score, all	Score, Female	Score, Male	1980-1990	1990-2000	2000-2010
(Intercept)	-11.5316^{***}	-11.4075^{***}	-11.5408^{***}	-10.9710^{***}	-11.1242^{***}	-11.3583^{***}
	(0.0103)	(0.0152)	(0.0127)	(0.0635)	(0.0317)	(0.0136)
SDI	-0.6974^{***}	-0.5629^{***}	-0.7256^{***}	-0.8111^{***}	-0.6738^{***}	-0.5693^{***}
	(0.0043)	(0.0067)	(0.0053)	(0.0236)	(0.0119)	(0.0056)
R^2	0.3107	0.3258	0.3060	0.2464	0.2550	0.3185
Adj. R^2	0.3107	0.3257	0.3060	0.2462	0.2550	0.3185
Num. obs.	57510	14562	42948	3613	9291	21983

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Score and SDI are in logs. SDI measure based on full (letter plus 2 digits of) JEL code.

Results: Broader diversification and rankings

Table 2: Diversification and performance: SDI with primary and secondary JEL codes

	Score, all	Score, Female	Score, Male	1980-1990	1990-2000	2000-2010
(Intercept)	-11.0739^{***}	-11.0513^{***}	-11.0499^{***}	-10.1632^{***}	-10.4503^{***}	-10.9322^{***}
	(0.0112)	(0.0166)	(0.0138)	(0.0651)	(0.0324)	(0.0146)
$SDI - L1$	-0.6484^{***}	-0.5169^{***}	-0.6721^{***}	-0.6745^{***}	-0.5503^{***}	-0.5065^{***}
	(0.0061)	(0.0095)	(0.0074)	(0.0324)	(0.0161)	(0.0078)
R^2	0.1669	0.1722	0.1633	0.1080	0.1121	0.1623
Adj. R^2	0.1669	0.1721	0.1633	0.1077	0.1120	0.1622
Num. obs.	56525	14291	42234	3583	9200	21661

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ Score and $SDI - L1$ are in logs. The SDI measure, $SDI - L1$ is based on the letter plus first digit of the JEL code.

Results: Broadest diversification and rankings

Table 3: Diversification and performance: SDI with primary JEL codes

	Score, all	Score, Female	Score, Male	1980-1990	1990-2000	2000-2010
(Intercept)	-10.4481^{***}	-10.6097^{***}	-10.3789^{***}	-9.2226^{***}	-9.7853^{***}	-10.4369^{***}
	(0.0116)	(0.0178)	(0.0142)	(0.0618)	(0.0306)	(0.0148)
LJEL1_HHI	-0.4513^{***}	-0.3834^{***}	-0.4608^{***}	-0.3063^{***}	-0.3315^{***}	-0.3581^{***}
	(0.0095)	(0.0149)	(0.0115)	(0.0485)	(0.0237)	(0.0119)
R^2	0.0398	0.0464	0.0380	0.0112	0.0213	0.0415
Adj. R^2	0.0397	0.0463	0.0380	0.0109	0.0212	0.0415
Num. obs.	54397	13707	40690	3511	8990	20986

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ Score and $SDI - L$ are in logs.
The SDI measure, $SDI - L$, uses only the letter part of the JEL code.

Robustness: Diversification and rankings, NEP

Table 4: Diversification and performance: NEP classification system

	Score, all	Score, Female	Score, Male	1980-1990	1990-2000	2000-2010
(Intercept)	-11.3118^{***}	-11.2877^{***}	-11.2917^{***}	-10.6015^{***}	-10.9273^{***}	-11.2063^{***}
	(0.0129)	(0.0189)	(0.0158)	(0.0658)	(0.0334)	(0.0155)
LNEP_HHI	-0.8576^{***}	-0.7138^{***}	-0.8873^{***}	-1.0059^{***}	-0.8642^{***}	-0.7280^{***}
	(0.0070)	(0.0106)	(0.0085)	(0.0334)	(0.0168)	(0.0083)
R^2	0.2448	0.2812	0.2384	0.2265	0.2466	0.2992
Adj. R^2	0.2448	0.2811	0.2384	0.2262	0.2465	0.2992
Num. obs.	46048	11488	34560	3100	8095	17872

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ Score and $SDI - L$ are in logs.

Table 5: The determinants of score

(Intercept)	-10.9103^{***}
	(0.0034)
LWScPages	0.0696^{***}
	(0.0016)
LWScCites	0.0585^{***}
	(0.0022)
NEPCites	0.0230^{***}
	(0.0002)
R^2	0.8073
Adj. R^2	0.8073
Num. obs.	45437

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Diversification –SDI– and the RePEc determinants of rankings

	LWScPages	LWScCites	NEPCites
(Intercept)	-0.1767^{***}	-0.4067^{***}	-10.0470^{***}
	(0.0345)	(0.0348)	(0.3705)
LJEL_HHI	-1.3083^{***}	-1.3452^{***}	-18.6688^{***}
	(0.0137)	(0.0138)	(0.1473)
R^2	0.1665	0.1723	0.2612
Adj. R^2	0.1665	0.1723	0.2612
Num. obs.	45437	45437	45437

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Co-authorship

Table 7: Co-authorship and performance

	Lscore	SDI	LWScPages	LWScCites	NEPCites	C per P
(Int)	−10.2464***	−2.1284***	2.1579***	1.8800***	22.2169***	3.9588***
	(0.0048)	(0.0041)	(0.0131)	(0.0129)	(0.1382)	(0.1928)
R_auth	0.0497***	−0.0264***	0.0804***	0.0935***	1.2123***	−0.0396***
	(0.0003)	(0.0003)	(0.0008)	(0.0008)	(0.0085)	(0.0119)
R^2	0.3885	0.1954	0.1793	0.2343	0.3096	0.0002
Adj. R^2	0.3885	0.1954	0.1793	0.2343	0.3095	0.0002
No.obs	45005	45005	45005	45005	45005	45005

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Early vs late diversification

Table 8: Number of distinct JEL codes in the first five years, the following ten years, and lifetime JEL Codes: Effect on Pages.

	LWScPages	LWScPages	LWScPages	LWScPages
(Intercept)	0.662^{***}	1.367^{***}	1.292^{***}	2.075^{***}
	(0.058)	(0.065)	(0.065)	(0.067)
LJEL_HHI	-1.124^{***}	-0.643^{***}	-0.733^{***}	-0.270^{***}
	(0.023)	(0.028)	(0.030)	(0.032)
JEL_0_5	-0.019^{***}		-0.020^{***}	-0.037^{***}
	(0.002)		(0.002)	(0.002)
JEL_5_10		0.028^{***}	0.029^{***}	
		(0.001)	(0.001)	
JEL_TOTAL				0.034^{***}
				(0.001)
R^2	0.104	0.117	0.121	0.161
Adj. R^2	0.104	0.117	0.121	0.161
Num. obs.	21620	21620	21620	21620

Early vs late diversification

Table 9: Number of different JEL codes in the first five years, the next ten years, and lifetime JEL Codes: Effects on Pages, Cohorts 2002-2004

	LWScPages	LWScPages	LWScPages	LWScPages
(Intercept)	0.807***	1.394***	1.448***	2.036***
	(0.140)	(0.157)	(0.158)	(0.163)
LJEL_HHI	−0.951***	−0.610***	−0.552***	−0.203**
	(0.057)	(0.069)	(0.074)	(0.078)
JEL_0_5	0.012*		0.013*	−0.017**
	(0.006)		(0.006)	(0.006)
JEL_5_10		0.027***	0.027***	
		(0.003)	(0.003)	
JEL_TOTAL				0.033***
				(0.002)
R^2	0.101	0.117	0.118	0.145
Adj. R^2	0.100	0.116	0.117	0.144
Num. obs.	3607	3607	3607	3607

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Early vs late diversification

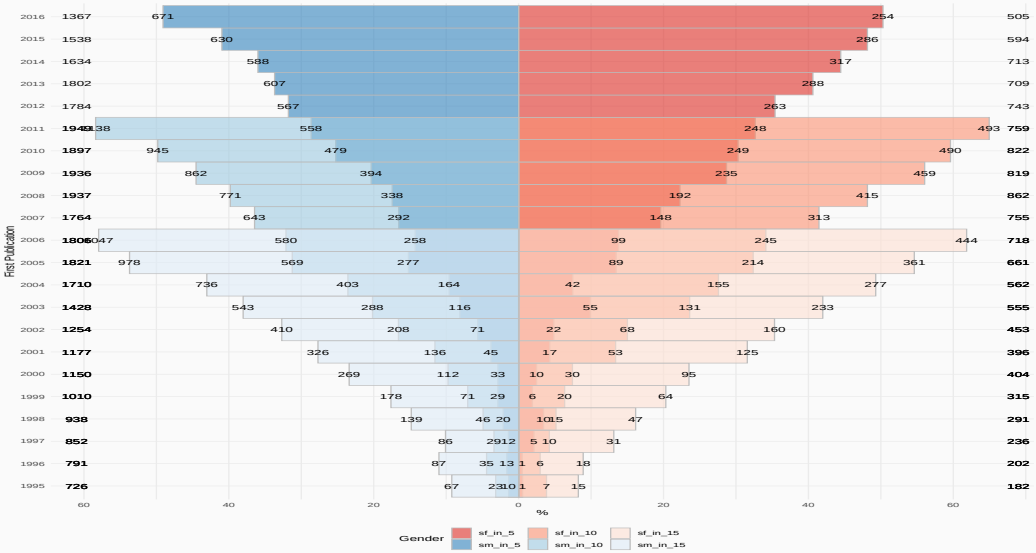
Table 10: Number of different JEL codes in the first five years, the next ten years, and lifetime JEL Codes and NEP Cites. Cohorts 2002-2004.

	NEPCites	NEPCites	NEPCites	NEPCites
(Intercept)	-3.254^{*}	7.427^{***}	9.170^{***}	15.666^{***}
	(1.613)	(1.784)	(1.797)	(1.848)
LJEL_HHI	-15.736^{***}	-9.894^{***}	-8.013^{***}	-4.235^{***}
	(0.655)	(0.785)	(0.837)	(0.880)
JEL_0_5	0.397^{***}		0.407^{***}	-0.045
	(0.067)		(0.065)	(0.068)
JEL_5_10		0.520^{***}	0.523^{***}	
		(0.037)	(0.037)	
JEL_TOTAL				0.513^{***}
				(0.028)
R^2	0.208	0.242	0.250	0.277
Adj. R^2	0.208	0.242	0.250	0.277
Num. obs.	3607	3607	3607	3607

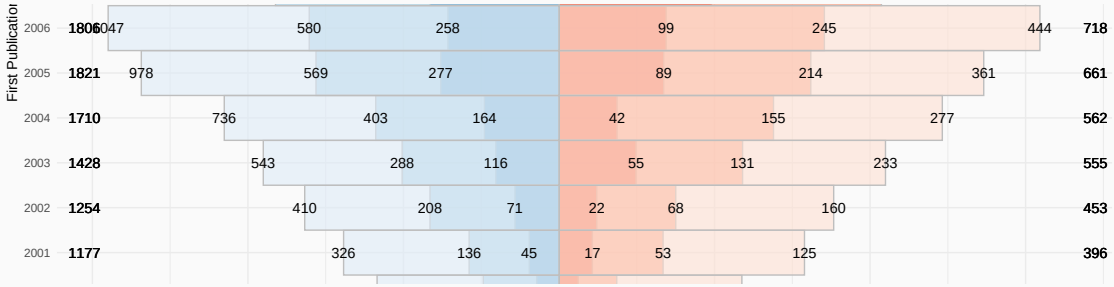
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Research attrition

Figure 5: Population Demographics.



Research attrition



Research attrition

Early vs late diversification

Table 11: Early diversification and attrition, Panel 2003-2012 cohorts

	Attr.Rate	Attr.Rate	Attr.Rate
(Intercept)	-0.5875^{***}	-0.4239^{***}	-0.1713^{***}
	(0.0392)	(0.0130)	(0.0242)
LJEL_HHI_5	0.4719^{***}		0.1606^{***}
	(0.0216)		(0.0132)
items_5		-0.0489^{***}	-0.0453^{***}
		(0.0013)	(0.0013)
AIC	34660.8109	32950.6269	32802.9424
BIC	34677.7699	32967.5859	32828.3809
Log Likelihood	−17328.4055	−16473.3134	−16398.4712
Deviance	34656.8109	32946.6269	32796.9424
Num. obs.	35579	35579	35579

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Early vs late diversification

Table 12: Early diversification and attrition- 2003-2004 cohorts

	Attr.Rate	Attr.Rate	Attr.Rate
(Intercept)	−0.2660*	−0.4044***	−0.0333
	(0.1164)	(0.0407)	(0.0713)
LJEL_HHI_5	0.7934***		0.2628***
	(0.0695)		(0.0427)
items_5		−0.0697***	−0.0606***
		(0.0049)	(0.0051)
AIC	3766.7687	3552.0276	3515.3309
BIC	3779.4889	3564.7478	3534.4111
Log Likelihood	−1881.3844	−1774.0138	−1754.6655
Deviance	3762.7687	3548.0276	3509.3309
Num. obs.	4273	4273	4273

Table 13: Lscore and SDI- 95th percentile of pages

Period	n	R-sq	adj. R-sq	LJEL_HHI
1981-1983	41	0.00	-0.02	-0.06
1984-1986	49	0.22	0.21	-0.76***
1987-1989	65	0.25	0.24	-0.98***
1990-1992	88	0.13	0.12	-0.51***
1993-1995	111	0.18	0.17	-0.60***
1996-1998	158	0.17	0.16	-0.64***
1999-2001	211	0.16	0.16	-0.55***
2002-2004	281	0.18	0.18	-0.55***
2005-2007	354	0.15	0.15	-0.58***
2008-2010	389	0.10	0.10	-0.37***
2011-2013	367	0.13	0.13	-0.35***
2014-2016	299	0.21	0.21	-0.44***

Table 14: Lscore and SDI- 95th percentile of pages

Period	n	R-sq	adj. R-sq	LJEL_HHI
1981-1983	41	0.00	-0.02	-0.06
1984-1986	49	0.22	0.21	-0.76***
1987-1989	65	0.25	0.24	-0.98***
1990-1992	88	0.13	0.12	-0.51***
1993-1995	111	0.18	0.17	-0.60***
1996-1998	158	0.17	0.16	-0.64***
1999-2001	211	0.16	0.16	-0.55***
2002-2004	281	0.18	0.18	-0.55***
2005-2007	354	0.15	0.15	-0.58***
2008-2010	389	0.10	0.10	-0.37***
2011-2013	367	0.13	0.13	-0.35***
2014-2016	299	0.21	0.21	-0.44***

Table 15: Lscore and SDI, Top 10 school graduates - 70-100th percentile of pages

Period	n	R-sq	adj. R-sq	LJEL_HHI
1981-1983	34	0.01	-0.02	-0.16
1984-1986	38	0.27	0.25	-0.73***
1987-1989	55	0.28	0.27	-1.05***
1990-1992	51	0.11	0.09	-0.49**
1993-1995	55	0.22	0.20	-0.65***
1996-1998	61	0.23	0.21	-0.76***
1999-2001	71	0.22	0.21	-0.59***
2002-2004	70	0.23	0.21	-0.74***
2005-2007	60	0.09	0.08	-0.46**
2008-2010	57	0.19	0.18	-0.61***
2011-2013	38	0.31	0.29	-0.59***
2014-2016	28	0.22	0.19	-0.48**

Table 16: Lscore, SDI, number of items- Top 10 school graduates, 70-100th of pages

Period	n	R-sq	adj. R-sq	LJEL_HHI	items
1981-1983	34	0.39	0.36	0.27	0.01***
1984-1986	38	0.43	0.40	-0.49**	0.01***
1987-1989	55	0.56	0.54	-0.31	0.01***
1990-1992	51	0.26	0.23	-0.33*	0.01***
1993-1995	55	0.45	0.43	-0.74***	0.01***
1996-1998	61	0.51	0.49	-0.61***	0.01***
1999-2001	71	0.38	0.36	-0.43***	0.01***
2002-2004	70	0.49	0.48	-0.51***	0.02***
2005-2007	60	0.33	0.30	-0.33*	0.02***
2008-2010	57	0.29	0.26	-0.47***	0.01**
2011-2013	38	0.48	0.45	-0.42***	0.03***
2014-2016	28	0.44	0.39	-0.30*	0.03***

Top 5 as a proxy of ability

Table 17: Distribution of Top 5.

Number of Top 5	*n*	Top5	*n*
1	2871	21	14
2	1186	22	11
3	631	23	11
4	353	24	10
5	303	25	9
6	234	26	4
7	164	27	7
8	114	28	3
9	96	29	3
10	79	30	5
11	58	31	7
12	66	32	4
13	43	33	3
14	38	34	5
15	34	35	2
16	31	36	1
17	28	37	2
18	25	38	6
19	20	39	2
20	14	40 +	16

Top 5 as a proxy of ability

Table 18: Score and top 5

	LScore	LScore	LScore	LScore	LScore	LScore
(Intercept)	-8.914^{***}	-10.263^{***}	-10.072^{***}	-11.462^{***}	-11.132^{***}	-11.099^{***}
	(0.012)	(0.020)	(0.017)	(0.027)	(0.025)	(0.025)
Grad_Rank	-0.002^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age		0.057^{***}	0.037^{***}	0.034^{***}	0.029^{***}	0.028^{***}
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
top5			0.148^{***}	0.132^{***}	0.119^{***}	0.118^{***}
			(0.002)	(0.002)	(0.002)	(0.002)
LJEL_HHI				-0.553^{***}	-0.332^{***}	-0.330^{***}
				(0.009)	(0.009)	(0.009)
Reg_authors					0.026^{***}	0.026^{***}
					(0.000)	(0.000)
genderb						-0.088^{***}
						(0.014)
R^2	0.114	0.390	0.577	0.666	0.743	0.744
Adj. R^2	0.113	0.390	0.577	0.666	0.743	0.744
Num. obs.	13691	13691	13691	13691	13691	13691

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Top 5 as a proxy of ability

Table 19: Pages and top 5

	LWScPag (3-4)	LWScPag (3-5)	LWScPag (3+)	LWScPag (4+)	LWScPag (6+)	LWScPag (8+)	LWScPag (10+)	LW
(Intercept)	5.301***	5.104***	5.810***	6.109***	6.455***	6.668***	6.836***	
	(0.140)	(0.095)	(0.058)	(0.065)	(0.077)	(0.094)	(0.103)	
top5	0.169***	0.223***	0.061***	0.053***	0.045***	0.038***	0.032***	
	(0.034)	(0.017)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
age	−0.006***	−0.005***	−0.005***	−0.007***	−0.010***	−0.010***	−0.010***	−
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
LJEL_HHI	−0.245***	−0.252***	−0.260***	−0.235***	−0.213***	−0.212***	−0.209***	−
	(0.025)	(0.021)	(0.016)	(0.017)	(0.019)	(0.023)	(0.024)	
R^2	0.149	0.230	0.499	0.505	0.532	0.529	0.545	
Adj. R^2	0.146	0.229	0.498	0.505	0.530	0.527	0.542	
Num. obs.	984	1287	2456	1825	1169	771	561	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: Academically older people without top 5 are low ranked

Top 5 as a proxy of ability

Table 20: Score, top 5, Age, and SDI (by groups of Top 5).

	LScore (3-4)	LScore (3-5)	LScore (3-6)	LScore (3-7)	LScore (3-8)	LScore (3-9)	LScore (10+)	LScore (15+)
(Intercept)	-10.814^{***}	-10.870^{***}	-10.736^{***}	-10.717^{***}	-10.658^{***}	-10.652^{***}	-9.400^{***}	-8.736^{***}
	(0.207)	(0.148)	(0.125)	(0.117)	(0.112)	(0.107)	(0.268)	(0.426)
top5	0.220^{***}	0.206^{***}	0.168^{***}	0.174^{***}	0.159^{***}	0.157^{***}	0.067^{***}	0.058^{***}
	(0.050)	(0.027)	(0.018)	(0.014)	(0.012)	(0.010)	(0.005)	(0.006)
age	0.016^{***}	0.016^{***}	0.015^{***}	0.014^{***}	0.014^{***}	0.013^{***}	0.007^{*}	-0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.005)
LJEL_HHI	-0.515^{***}	-0.553^{***}	-0.556^{***}	-0.553^{***}	-0.558^{***}	-0.561^{***}	-0.564^{***}	-0.550^{***}
	(0.037)	(0.033)	(0.031)	(0.029)	(0.029)	(0.028)	(0.063)	(0.096)
R^2	0.205	0.236	0.248	0.264	0.272	0.295	0.448	0.410
Adj. R^2	0.202	0.234	0.247	0.263	0.271	0.294	0.445	0.403
Num. obs.	984	1287	1521	1685	1799	1895	561	277

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$