

The impact of Grey Literature in the web environment: A citation analysis using Google Scholar

Rosa Di Cesare, Daniela Luzi, Roberta Ruggieri
Consiglio Nazionale delle Ricerche,
Istituto di Ricerche sulla Popolazione
e le Politiche Sociali

Mail to {r.dicesare, d.luzi, r.ruggieri}@irpps.cnr.it

Background of the study

Citation analysis applied to GL:

- citation counts in bibliographic references of conventional journal articles
- citation counts using ISI Web of Science and its bibliometric indicators

2004 development of new citation tracking systems:

Google Scholar

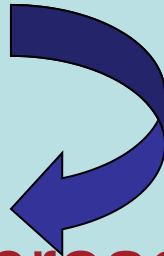


“Index of peer-reviewed papers, theses, preprints, and technical reports from all disciplines”

Objectives

Verify:

- Google Scholar (GS) is an efficient tool to:
 - *Identify core papers*
 - *Track citations from different types of document*
- GL documents receive citations
- The role of GL as citing documents



Is GS able to represent the scholarly communication deriving from citations considering both GL and conventional literature?

Methodology

Cited documents

Step 1: Querying “population ageing” as exact phrase in the title.

Result: 1419 documents

Step 2: Selection of the documents with at least **10 citations**.

Result: 99 documents → **Cited documents**

Step 3: Analysis of the 99 cited documents:

- publication type
- document type
- publication year
- availability of full text
- topic

Step 4: Selection of the cited documents with at least **50 citations**.

Result 15 documents → **Highly cited documents**

Methodology 2)

Documents citing highly cited documents

Step 5: Identification of citing documents through data validation to exclude duplicates, missing references and incomplete bibliographic elements.

Full text retrieval essential to check bibliographic data

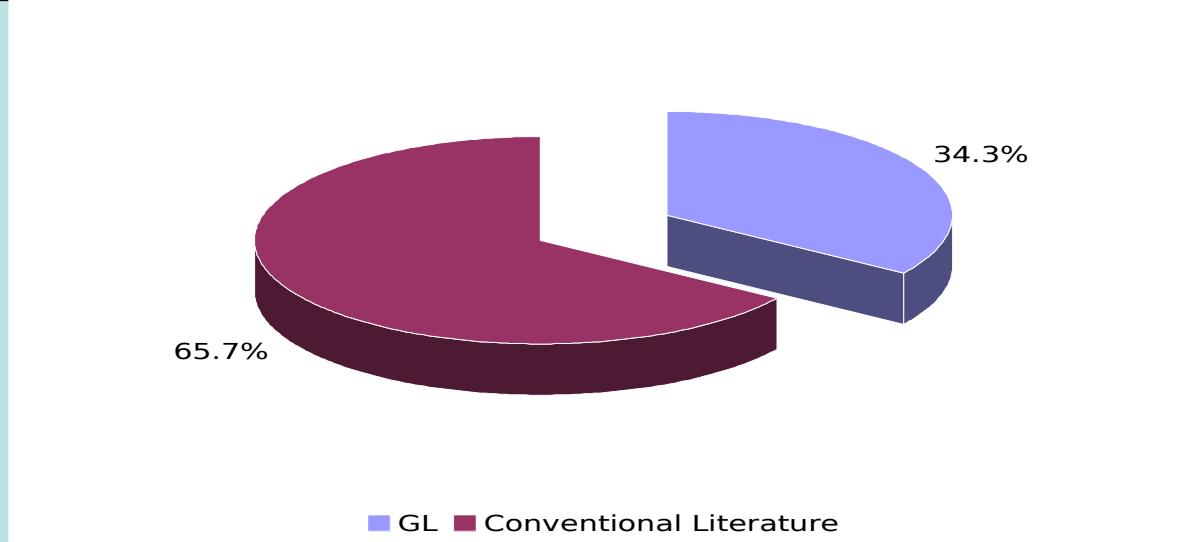
Result 885 documents

Step 6: Analysis of the citing documents:

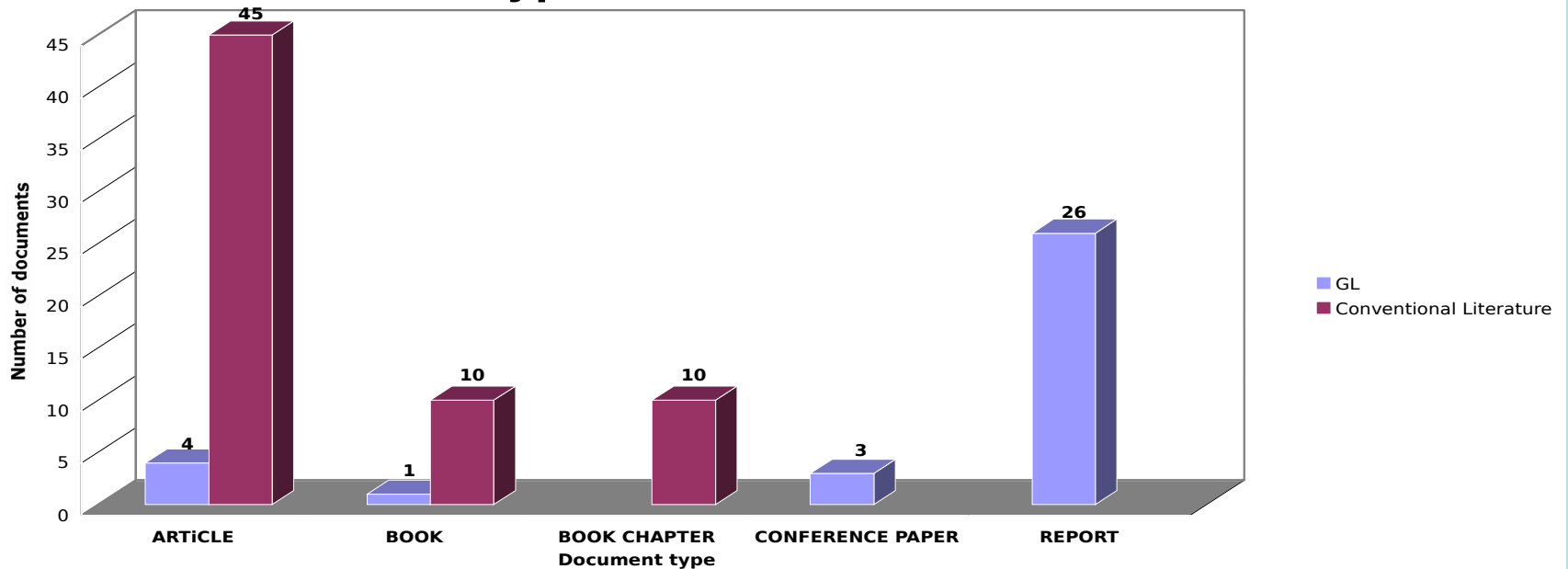
- publication type
- document type
- date of publication
- availability of full text
- topic
- self citation
- language

Are GL documents ever cited?

Does GS give visibility to GL documents? (n=99)

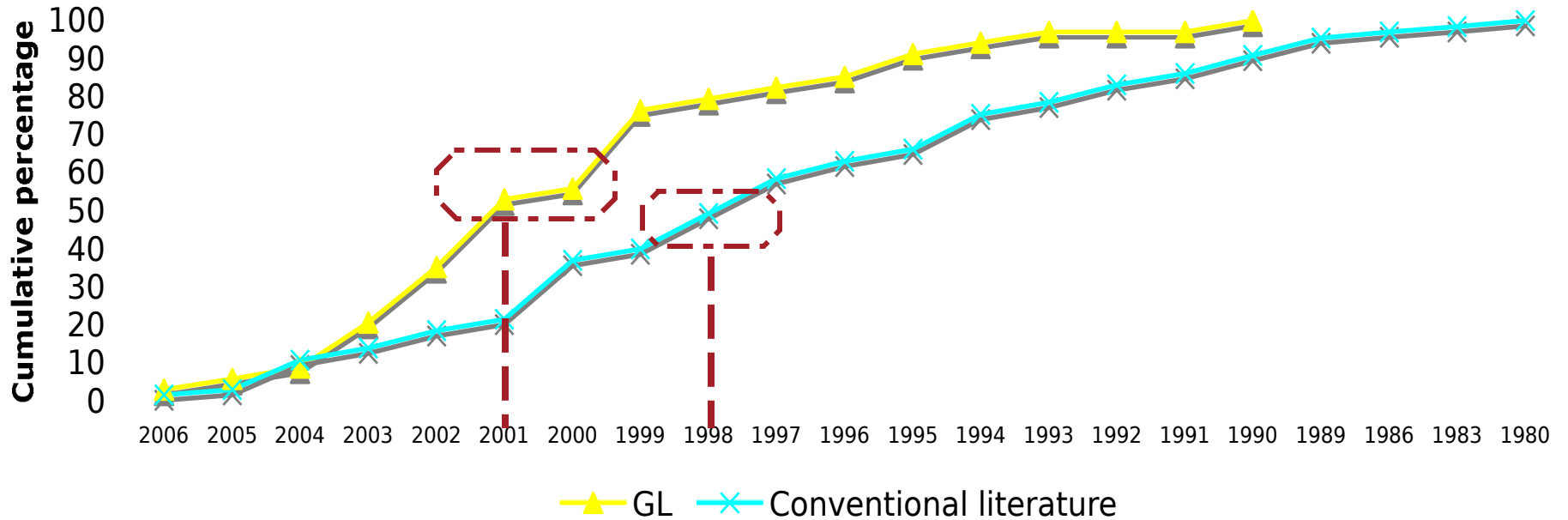


Which types of documents are cited?



Are GL and conventional documents concentrated in different specific time interval? (n=99)

Distribution of cited documents by publication year



Data validation of the citing documents the 15 highly cited documents

Papers	Type of document cited	GS citing documents	Duplicates		Missing reference		Incomplete bibliographic elements		Citations analysed
			No.	%	No.	%	No.	%	
P1	ARTICLE	158	6	3.8	--	--	5	3.2	147
P2	REPORT	105	11	10.5	2	1.9	6	5.7	86
P3	REPORT	101	7	6.9	3	3.0	8	7.9	83
P4	ARTICLE	78	5	6.4	--	--	8	10.3	65
P5	ARTICLE	72	6	8.3	--	--	3	4.2	63
P6	REPORT	65	11	16.9	--	--	6	9.2	48
P7	REPORT	63	7	11.1	--	--	6	9.5	50
P8	ARTICLE	62	2	3.2	2	3.2	5	8.1	53
P9	ARTICLE	61	8	13.1	4	6.6	5	8.2	44
P10	ARTICLE	55	7	12.7	2	3.6	7	12.7	39
P11	REPORT	54	11	20.4	3	5.6	--	--	40
P12	REPORT	55	7	12.7	2	3.6	7	12.7	39
P13	ARTICLE	53	4	7.5	5	9.4	3	5.7	41
P14	REPORT	53	2	3.8	2	3.8	4	7.5	45
P15	REPORT	53	3	5.7	6	11.3	2	3.8	42
Total		1088	97	8.9	31	2.8	75	6.9	885

81% of valid data
19% discarded

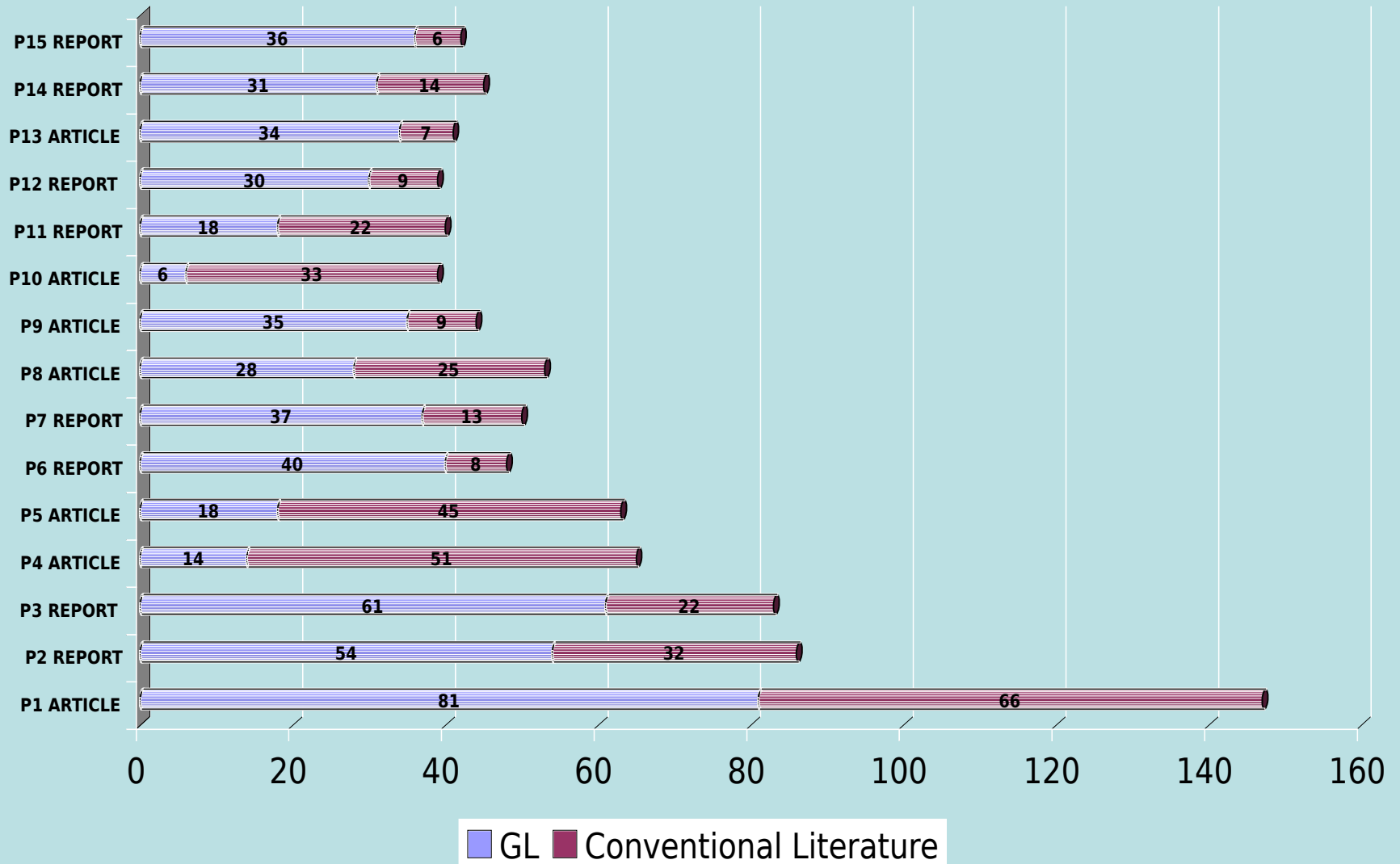
Profile of the 15 highly cited documents

Highly cited documents	Reports		Journal articles		Total	
<i>Citation range</i>	No.	%	No.	%	No.	%
≥80	2	13.3	1	6.7	3	20.0
79-50	1	6.7	3	20.0	4	26.7
< 50	5	33.3	3	20.0	8	53.3
Τοιολ	8	53.3	7	46.7	15	100.0
Δατε φ πυβλχαιον						
1990-1994	2	13.3	3	20.0	5	33.3
1995-1999	3	20.0	2	13.3	5	33.3
2000-2003	3	20.0	2	13.3	5	33.3
Χιταιον ιμπαχι						
Ιν τησσιμε ψεορ οφτυβλχαιον	5	33.3	3	20.0	8	53.3
Ιν στηερεαφροσ	3	20.0	4	26.7	7	46.7
Χιταδ ιν2007	5	33.3	5	33.3	10	66.7
Χιταδ ινοτηφ ψεαροσ	3	20.0	2	13.3	5	33.3
Φυλλεξτ						
Απαλαβε	6	40.0	2	13.3	8	53.3
Νοτ παλβλε	2	13.3	5	33.3	7	46.7
Τοπιχ						
Ηεολτη αεχισ	2	13.3	3	20.0	5	33.3
Εχονομιχ αεχισ	5	33.3	2	13.3	7	46.7
Δεμο-σχιολ αεχισ	1	6.7	2	13.3	3	20.0

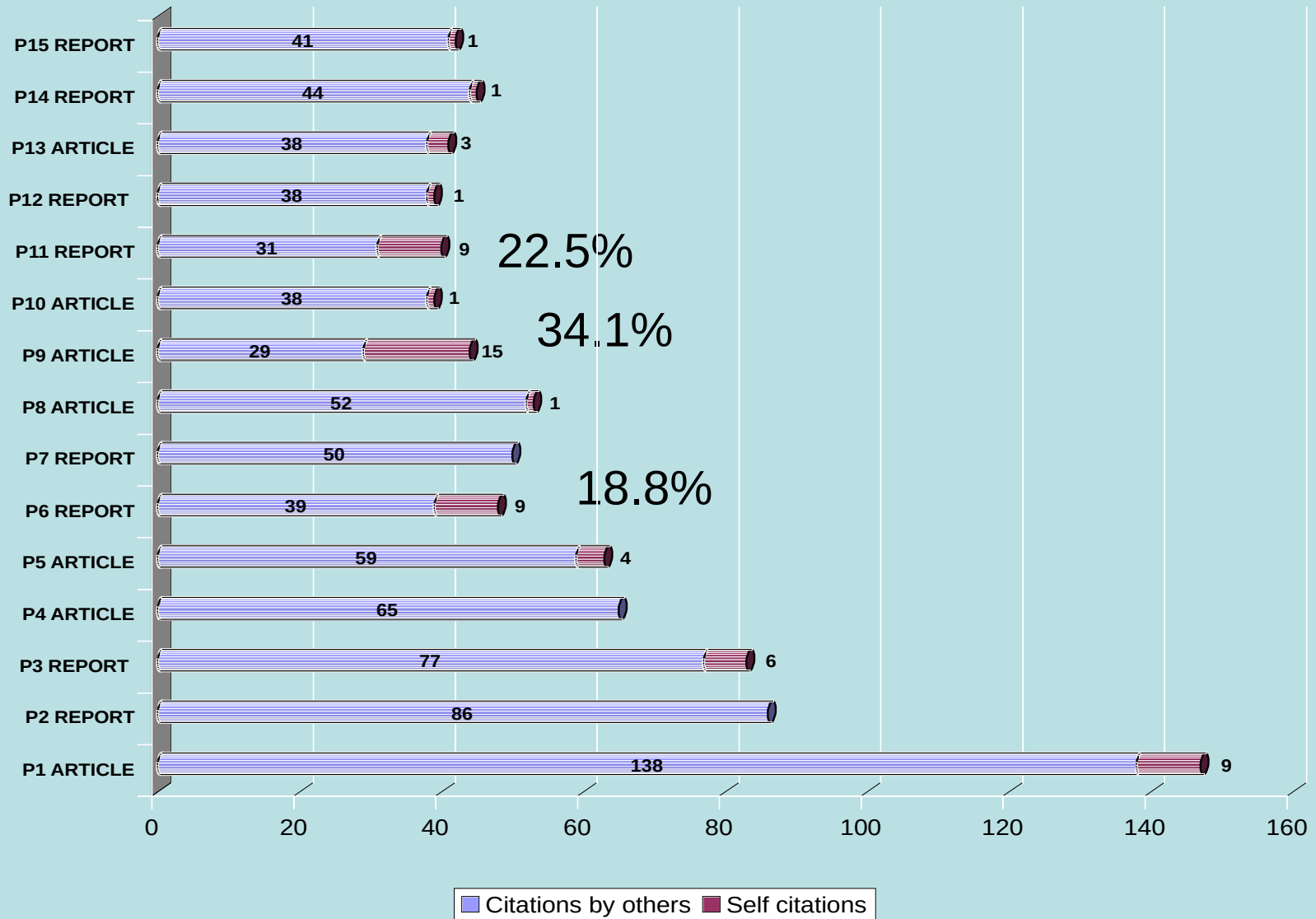
Profile of the 885 citing documents

Characteristics of citing documents	All citing documents		Publication type			
	All citing documents		Grey literature		Conventional literature	
	No.	%	No.	%	No.	%
	885	Total	523	59.1	362	40.9
Documents type	No.	%	No.	%	No.	%
Journal article	416	47.0	72	13.8	344	95.0
Report	302	34.1	302	57.7	--	--
Conference paper	85	9.6	85	16.3	5	1.4
Thesis	33	3.7	33	6.3	--	--
Draft	21	2.4	21	4.0	--	--
Book and book chapter	28	3.2	10	1.9	13	3.6
Age	No.	%	No.	%	No.	%
Less than 5 years	426	48.1	253	48.4	173	47.8
5 years - 10 years	403	45.5	234	44.7	169	46.7
More than ten years	31	3.5	17	3.3	14	3.9
Not available	25	2.8	19	3.6	6	1.7
Citation type	No.	%	No.	%	No.	%
Self-citation	60	6.8	33	6.3	27	7.5
Citation by other	825	93.2	490	93.7	335	92.5
Language	No.	%	No.	%	No.	%
English	762	86.1	426	81.5	336	92.8
# English	123	13.9	97	18.5	26	7.2
Full text	No.	%	No.	%	No.	%
Available	577	65.2	499	95.4	78	21.5
Not available	308	34.8	24	4.6	284	78.5

Number of citations given to the highly cited documents by publication type



Number of citations by self-citations and citations by others



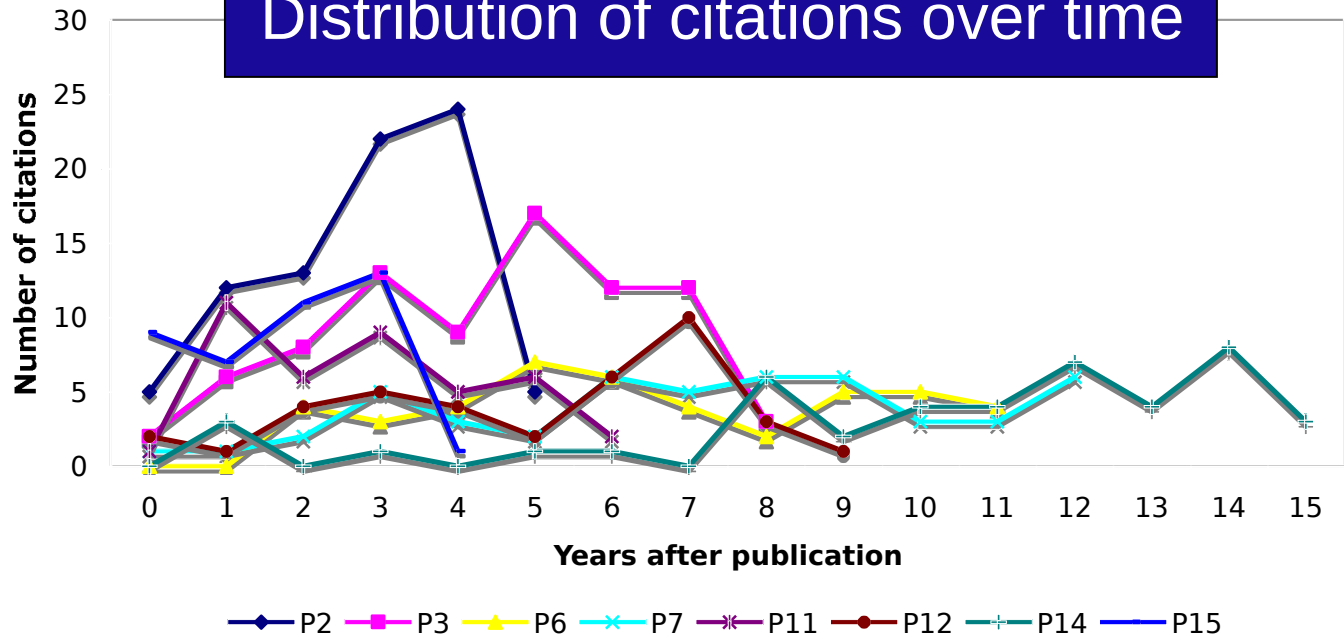
Important figures of citations received by the 15 highly cited documents

Papers	Document type	Publication year	Total citations	FCY	LCY	TCY	Citat	Peak
							Average/year	
P1	ARTICLE	1999	147	1999	2007	9	16.3	32 (2006)
P2	REPORT	2002	86	2002	2007	6	14.3	24 (2006)
P3	REPORT	1999	83	1999	2007	9	9.2	17 (2004)
P4	ARTICLE	2000	65	2000	2007	8	8.1	14 (2003)
P5	ARTICLE	1992	63	1995	2007	13	4.8	8 (2004)
P6	REPORT	1995	48	1997	2006	10	4.8	7 (2000)
P7	REPORT	1994	50	1994	2006	13	3.8	6 (2006)
P8	ARTICLE	1994	53	1994	2006	13	4.0	11 (2006)
P9	ARTICLE	1992	44	1997	2006	10	4.4	7 (2003)
P10	ARTICLE	1999	39	2000	2007	8	4.8	12 (2003)
P11	REPORT	2001	40	2001	2007	7	5.7	11 (2002)
P12	REPORT	1998	39	1998	2007	10	3.9	10 (2005)
P13	ARTICLE	2001	41	2002	2007	6	6.8	11 (2005)
P14	REPORT	1990	45	1991	2005	12	3.7	8 (2004)
P15	REPORT	2003	42	2003	2007	7	6.0	13 (2006)

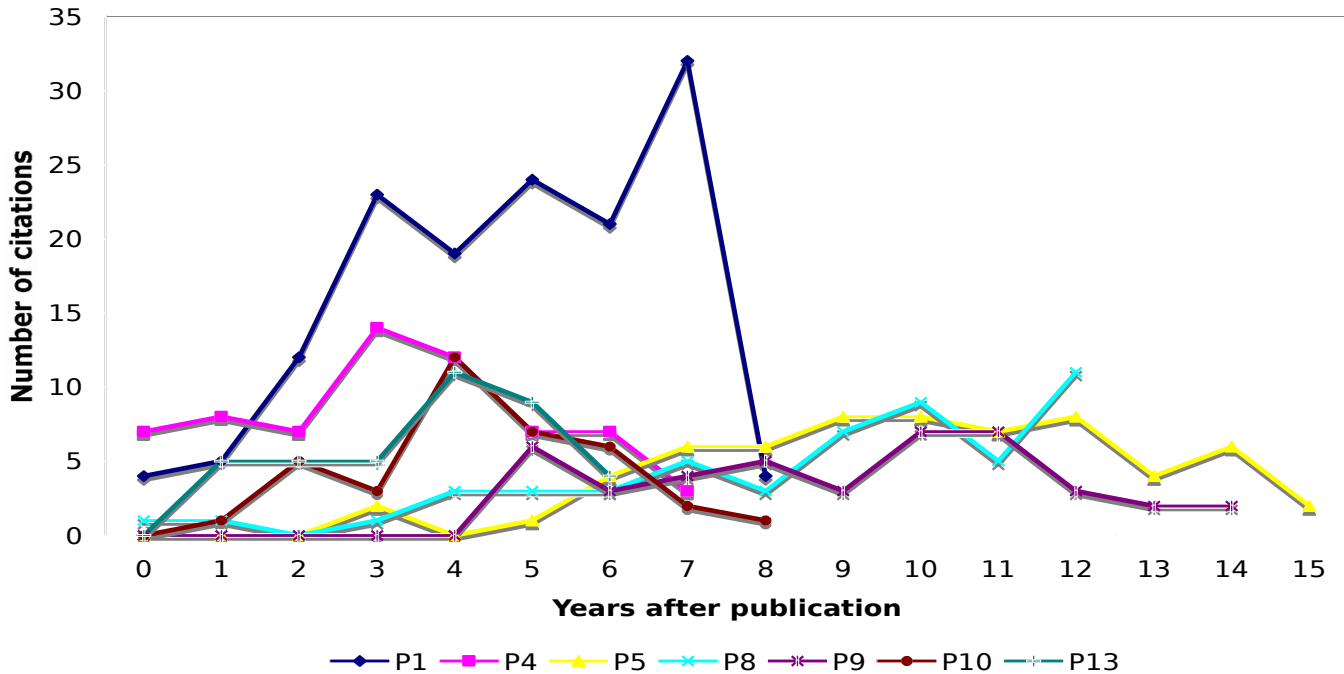
FCY=First citation year; LCY= Last citation year; TCY=Total citation years

GL

Distribution of citations over time



Conventional literature



Conclusions

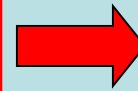
Google Scholar

- visibility of **GL** both as **cited** and **citing** documents
- GL is **evidently** part of core papers
- possibility of reconstructing the **document life-cycle**: from GL to conventional literature

Citation analysis

Citation counts, citations over time, immediacy impact, self-citations

- Highly cited GL receives more citations by GL documents and *viceversa*
- GL receives a higher number of citations in the first 5 years
- Language



Differences between GL and conventional literature tend to disappear



Some typical characteristics of GL can be still noticed