

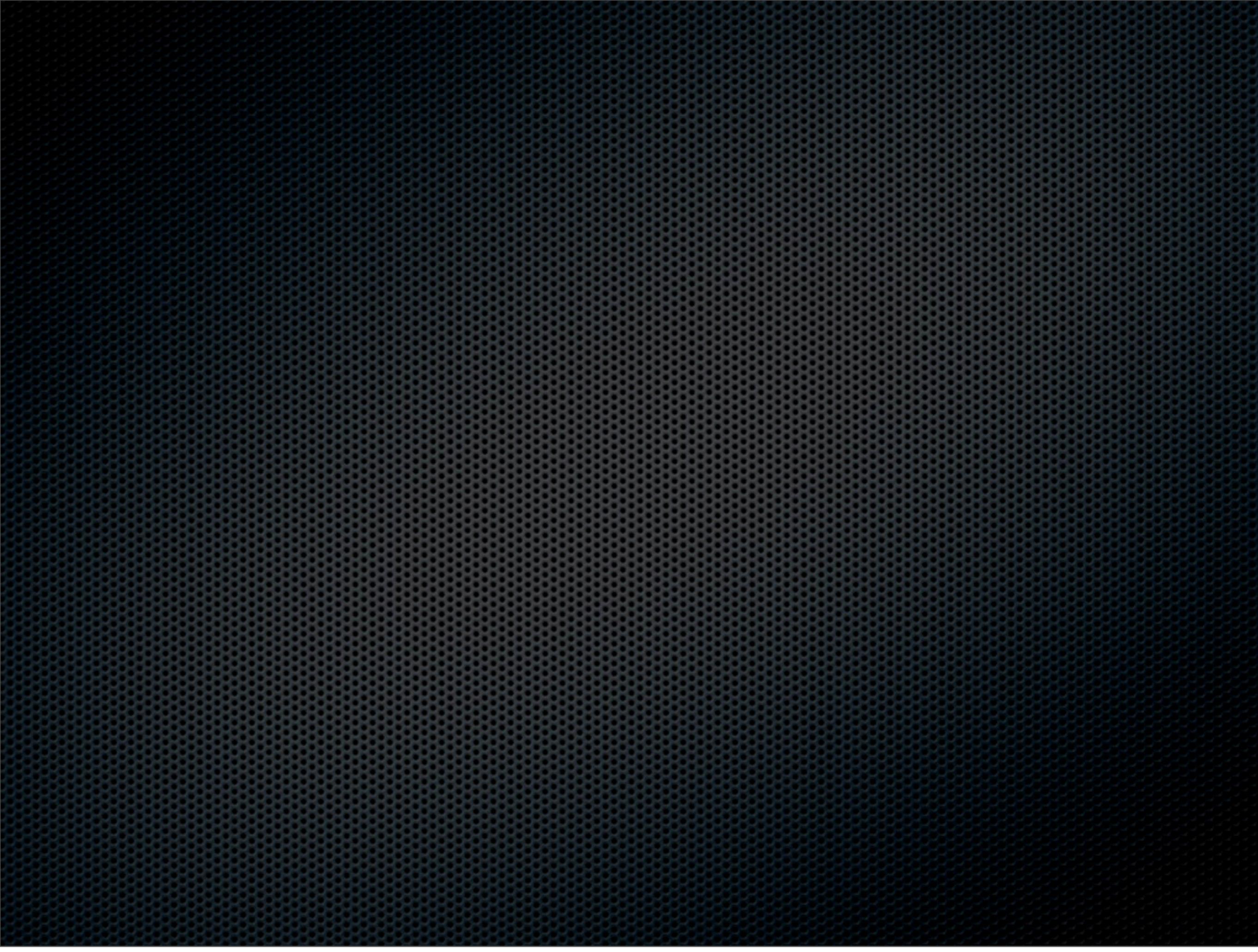
# How Scholarly is Google Scholar<sup>TM</sup>?

## A Comparison of Google Scholar to Library Databases

Jared L. Howland | Thomas C. Wright | Rebecca A. Boughan | Brian C. Roberts  
Brigham Young University

Introduction

# Literature Review



1998  
NBASearch: [Custom Search](#)[Dogpile Remote](#)[Search at Home](#)[Barnes & Noble](#)  
[Computer](#)[UPS](#)  
[Road](#)[WEB](#)  
[Auto](#)  
[Business](#)[Careers](#)  
[Computers](#)  
[Education](#)[Entertainment](#)  
[Fashion](#)  
[Games](#)[Government](#)  
[Health](#)  
[Kids](#)[About Lycos](#)  
[Business Dev](#)[Link to Lycos](#)  
[Copyright](#) (c) 1998 Lycos

All Rights Reserved.

Lycos (r) is a register  
of Carnegie Mellon Univ



Search the web using Google!

10 results



Google Search

I'm feeling lucky

*Index contains ~25 million pages (soon to be much bigger)*

## About Google!

[Stanford Search](#) [Linux Search](#)

Get Google! updates monthly!

your e-mail

Subscribe

[Archive](#)

Copyright ©1997-8 Stanford University

Methodology

## Subject specialists

7 Subject specialists from 3 disciplines:

- 3 from sciences
- 2 from humanities
- 2 from social sciences

Blind to purpose of study

Asked them to give us 3 things...

# How does the acquisition and use of a second language in children affect their general cognitive development?

A question they would typically receive from a student  
(humanities)

(bilingual\* OR L2) AND  
(child\* OR toddler) AND  
“cognitive development”

A structured query they would use to search a database

# Linguistic and Language Behavior Abstracts

The database they would use to search for that question

Academic Discipline	Database Query	Library Database
Science	(ACL or "anterior cruciate ligament") and injur* and (athlet* or sport or sports) and (therap* or treat* or rehab*)	SportDiscus
Science	lung cancer and (etiol* or caus*) and (cigarette* or smok* or nicotine*)	Medline
Science	“dark matter” and evidence	Applied Science and Technology Abstracts
Social Science	(“fast food” or mcdonald’s or wendy’s or “burger king” or restaurant) and franchis* and (knowledge n3 transfer or “knowledge management” or train*)	Business Source Premier
Social Science	(“standardized test*” or “high stakes test*”) and (“learning disabilit*” or Dyslexia or “learning problem”) and accommodat*	PsycINFO
Humanities	(bilingual* or L2) and (child* or toddler) and “cognitive development”	Linguistics and Language Behavior Abstracts
Humanities	(memor* or remembrance or memoir*) and (holocaust) and (Spiegelman or Maus)	JSTOR

This is what things looked like after we got all the information back from the librarians

Methodology

## Search using query

Then we took that information and used it in 2 ways.

The screenshot shows the CSA Illumina search results page. At the top, there's a blue header bar with the CSA logo, the word 'ILLUMINA', and a sub-headline 'Is a new Cold War with Russia heating up?'. Below the header are navigation links: Logout, Quick Search (highlighted in blue), Advanced Search, Search Tools, and Browse. A 'Results' section follows, displaying '146 results found for: (bilingual\* or L2) and (child\* or toddler) and "cognitive development ... in Multiple Data'.

Below the search bar, there are several categories with counts: Published Works 146, Scholars 120, Web Sites 1, All Publication Types 146, Journals 110, Peer-Reviewed Journals 55, Conferences 3, Books 11. Below these are links for 'Mark or Clear all on page', 'Update Marked List', 'Save, Print, Email', and 'RefWorks'.

The main content area displays two search results:

1. [The Universality of Symbolic Representation for Reading in Asian and Alphabetic Languages](#)  
Bialystok, Ellen; Luk, Gigi  
*Bilingualism: Language and Cognition*, 2007, 10, 2, July, 121-129  
... understanding of the symbolic function of print, is common to **children** learning to read in these two different systems. Four-year-old **children** in Hong Kong learning to read in Cantonese & **children** in Canada learning to read ...  
[View Record](#) | [GET IT! @ BYU](#)
2. ["When you write 'four' in Chinese, you will find two 'J's' in it": A case study of four children learning to be literate in alphabetic and non-alphabetic print](#)  
Lu, Mei-Yu  
*Dissertation Abstracts International, A: The Humanities and Social Sciences*, 2007, 67, 08, Feb, 2914  
... qualitative case study approach to examine the nature of **bilingual** and biliteracy learning process of four young ethnic Chinese **children** living in a community where mainstream American culture and English predominated. I used observations, ...  
[View Record](#) | [GET IT! @ BYU](#)

At the bottom right of the results area, there are navigation links: 'Previous' and page numbers 1, 2, 3, 4, 5, 'Next'.

# Native database results

The first was to actually run the search query in the suggested database.

We put the first 30 citations into a bibliographic citation manager and saved all of the actual full text

We chose 30 because usability studies (Jakob Nielsen) tell us that less than 1% of all users ever go beyond the 3rd page of results and very few people ever change the defaults (ie, once they run a search they stick with it, success or failure).

Most of our DBs present 10 results per page so 30 results should represent a large enough sample to represent the actual set of results the majority of our users is ever going to see after performing a search.

The screenshot shows a Google Scholar search results page. The search query is `~bilingual OR L2 ~child OR toddler *cognitive developmen`. The results are filtered to show **Scholar All articles - Recent articles**, with a total of **Results 1 - 10 of about 19,3**. The first result is a link to an article titled **The influence of bilingualism on cognitive strategy and cognitive development**, with **all 5** versions available. The second result is **Bilingual cognitive development: Addressing three gaps in current research**, also with **all 3** versions available. The third result is **Bilingualism and cognitive development**, with **all 3 versions** available. The fourth result is a book titled **Language Acquisition and Conceptual Development**, with **all 5 versions** available. The fifth result is **Cognitive Complexity and Attentional Control in the Bilingual Mind**.

# Google Scholar results

We ran the same query in Google Scholar and saved the results again in a bibliographic Manager.

We used Zotero to quickly export all of the results.

We also saved the full text of each citation for later use in our study.

Methodology

## Search using citations

So, the first searches we ran using the native DBs and GS was for the query given to us by the librarian

The second set of searches we ran was to see if the citations we found in the DB were available in GS and vice versa

The screenshot shows a search results page from CSA Illumina. The search query was '(bilingual\* or L2) and (child\* or toddler) and "cognitive development ... in Multiple Data'. The results section displays 146 items, with 120 being scholars and 1 being a web site. The first two results are highlighted with a red oval:

- 1. The Universality of Symbolic Representation for Reading in Asian and Alphabetic Languages**  
Bialystok, Ellen; Luk, Gigi  
*Bilingualism: Language and Cognition*, 2007, 10, 2, July, 121-129  
... understanding of the symbolic function of print, is common to **children** learning to read in these two different systems. Four-year-old **children** in Hong Kong learning to read in Cantonese & **children** in Canada learning to read ...  
[View Record](#) | [GET IT! @ BYU](#)
- 2. "When you write 'four' in Chinese, you will find two 'J's' in it": A case study of four children learning to be literate in alphabetic and non-alphabetic print**  
Lu, Mei-Yu  
*Dissertation Abstracts International, A: The Humanities and Social Sciences*, 2007, 67, 08, Feb, 2914  
... qualitative case study approach to examine the nature of **bilingual** and biliteracy learning process of four young ethnic Chinese **children** living in a community where mainstream American culture and English predominated. I used observations, ...  
[View Record](#) | [GET IT! @ BYU](#)

# Is this citation available in Google Scholar?

Here is the same screenshot we saw just a minute ago.

We took the bibliographic information for each citation and searched for the citation within Google Scholar.



[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

the universality of symbolic representation for reading

[Advanced Scholar Search](#)  
[Scholar Preferences](#)  
[Scholar Help](#)

The following words are very common and were not included in your search: the of for in. [\[details\]](#)  
The "AND" operator is unnecessary -- we include all search terms by default. [\[details\]](#)

**Scholar All articles - Recent articles** Results 1 - 10 of about 8,790 for the [universality of symbolic representation for reading](#)

**[The universality of symbolic representation for reading in Asian and alphabetic languages - GET IT! @ BYU](#)**

E BIALYSTOK, G LUK - Bilingualism: Language and Cognition, 2007 - Cambridge Univ Press  
... 121 The **universality of symbolic representation for reading in Asian and alphabetic languages** \* ELLEN BIALYSTOK GIGI LUK York University ...

[Cited by 1](#) - [Related Articles](#) - [Web Search](#) - [BL Direct](#)

**[Brain strategies for reading in the second language are determined by the first language - all 3 versions »](#)**

T Nakada, Y Fujii, IL Kwee - Neuroscience Research, 2001 - Elsevier  
... the LG are not the **universal** neuroanatomic substrate ... the complexity of the written **symbolic representation** to which ... in subjects whose first **reading** system is ...

[Cited by 24](#) - [Related Articles](#) - [Web Search](#) - [Check BYU](#)

# Yes, it is available

We then did the same thing in reverse.

We took the 30 results from GS and searched for each citation within the database



## Exclusivity

This allowed us to later calculate something we called “exclusivity”

We put the citations into 1 of 3 possible “exclusivity” categories

Shows proportion of citations within our study that overlap. As you can see, within our study we found that, on average, GS had a larger result set overall as well as more exclusively than the databases.

Methodology

## Citation grading

So now that we have the citations from the database and the citations from Google Scholar. We used the bibliographic manager to generate a list of references that we input into an Excel spreadsheet. Then, using a random number table, we completely randomized the order of the citations for each subject specialist.

Citation Number	References
<a href="#">01</a>	Caporali, S. A., de Lacerda, C. B. F., & Marques, P. L. (2005). Teaching sign language to the families of the deaf: Focusing on the learning process. <i>PRO-FONO: Revista de Atualizacao Cientifica</i> , 17(1), 89-98.
<a href="#">02</a>	Francis, N. (2002). Modular perspectives on bilingualism. <i>International Journal of Bilingual Education and Bilingualism</i> , 5(3), 141-161.
<a href="#">03</a>	Chan, K. T. (2004). Chinese-English bilinguals' theory-of-mind development (Doctoral dissertation, University of Toronto, 2004).
<a href="#">04</a>	Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. <i>TESOL Quarterly</i> , 26(2), 319-343.
<a href="#">05</a>	Auestad, N., Scott, D. T., Janowsky, J. S., Jacobsen, C., Carroll, R. E., Montaldo, M. B., et al. (2003). Visual, cognitive, and language assessments at 39 months: A follow-up study of children fed formulas containing long-chain polyunsaturated fatty acids to 1 year of age. <i>Pediatrics</i> , 112(3), e177-e183.
<a href="#">06</a>	Colson, E. R., & Dworkin, P. H. (1997). Toddler development. <i>Pediatrics in Review</i> , 18(8), 255-259.
<a href="#">07</a>	Hendriks, H., Watorek, M., & Giuliano, P. (2004). The expression of localization and movement in descriptions and narratives in L1 and L2. <i>Langages</i> , 155(Sept), 106-126.
<a href="#">08</a>	Andrews, D. B. (2004). The acquisition of Spanish gender by English-speaking children in a partial immersion setting (Doctoral dissertation, University of Washington, 2004).
<a href="#">09</a>	Cable, C. (2004). 'I'm going to bring my sense of identity to this': The role and contribution of bilingual teaching assistants. <i>Westminster Studies in Education</i> , 27(2), 207-222.
<a href="#">10</a>	Long, M. (2005). Problems with supposed counter-evidence to the critical period hypothesis. <i>Iral</i> , 43(4), 287-317.

Finally, to deliver the content to the librarians in a way in which it would be easiest for them to evaluate, we saved the full-text of each citation according to its randomly assigned citation number. Then we used Excel to create hyperlinks to the full-text of each citation and delivered this list along with the full-text on a CD to the subject librarians. We asked them to evaluate each citation using a rubric which we provided in hard copy form. As you can see, the subject librarians were only able to see the citation number and the bibliographic information. By clicking on the hyperlinked citation number, the full-text of that citation would appear and the subject librarians could easily rate the citation on the rubric.

Have full text appear on this page after click to simulate linking from provided document.

Citation Number	References
<a href="#">01</a>	Caporali, S. A. Focusing on t
<a href="#">02</a>	Francis, N. (2005). <i>Bilingualism</i> , 5(1), 1-18.
<a href="#">03</a>	Chan, K. T. (2004). <i>Modular Perspectives on Bilingualism</i> . Toronto, 2004.
<a href="#">04</a>	Block, E. L. (1997). <i>TESOL Quarterly</i> , 32(2), 319-343.
<a href="#">05</a>	Auestad, N., S. (2003). Visual, cognitive, and chain polyuns
<a href="#">06</a>	Colson, E. R., and narrative
<a href="#">07</a>	Hendriks, H., and narrative
<a href="#">08</a>	Andrews, D. B. setting (Doctoral dissertation)
<a href="#">09</a>	Cable, C. (2003). assistants. <i>W</i>
<a href="#">10</a>	Long, M. (2003). <i>Modular Perspectives on Bilingualism</i> . 317.

## Modular Perspectives on Bilingualism<sup>1</sup>

*Norbert Francis*

*Northern Arizona University, Education - Box 5774, Northern Arizona University, Flagstaff, Arizona 86011, USA*

The following review of the research traces the current discussion on models of bilingualism to the contributions of L. Vygotsky and A.R. Luria. The attempt to differentiate among the different components of language ability in child development has from the beginning sparked sharp debate. Bilingualism, as an object of study, offers researchers a privileged viewpoint on the questions in dispute: the distinction between interpersonal conversational discourse and literacy-related academic discourse, the relationship between the development of linguistic knowledge of two languages and general cognitive development, and aspects of development that are universal and those aspects of child bilingual development that show wide variation. The article proposes that a modular approach to studying the different aspects of bilingual development promises to chart a course toward finding a broader common ground around research findings and interpretations that currently appear to be irreconcilable.

### Introduction

In the study of bilingualism, research is drawn toward multidisciplinary approaches to a greater degree than in the linguistic and cognitive sciences in general. Tied closely as it has been to applied fields in bilingual education and second language learning, researchers borrow freely from theoretical and experimental work in a number of areas. Since bilingualism involves another layer of interactions among the different cognitive domains, it has been difficult to delimit the investigation to one or another. This of course both offers opportunities and opens up many traps to fall into. On balance, however, the interchange (mostly one-sided for now) between the applied-oriented and theoretically-oriented research programmes is probably a good thing, not just for the former. This discussion paper will explore one such application of a proposal from linguistics and cognitive science: how the concept of modularity, or cognitive models of the modular type, may help bring clarity to a number of vexing conceptual sticking points that have persisted in the research. In particular, the inquiry and debate that interests us here concerns the relationship between linguistic knowledge and the competencies that underlie what have variously been described as the higher-order discourse abilities and proficiencies related to literacy and academic language use, competencies that are associated with secondary discourses learned through schooling, and its extracurricular counterparts.

In accord with our multiple focus, we will consider what appear to be some historical antecedents to the modular approach (from outside its traditional sphere of investigation), apply it to a current leading model of bilingual proficiency (that of J. Cummins), and assess a recent critique of Cummins' model (MacSwan, 2000) distinguished as it is from the more well trodden line of criti-

) the families of the deaf: -98.

*al Education and*

*insertion, University of*

*rs. *TESOL Quarterly*, 26*

: al. (2003). Visual, formulas containing long-

, 255-259.

vement in descriptions

in a partial immersion

tion of bilingual teaching

hesis. *Iral*, 43(4), 287-

Finally, to deliver the content to the librarians in a way in which it would be easiest for them to evaluate, we saved the full-text of each citation according to its randomly assigned citation number. Then we used Excel to create hyperlinks to the full-text of each citation and delivered this list along with the full-text on a CD to the subject librarians. We asked them to evaluate each citation using a rubric which we provided in hard copy form. As you can see, the subject librarians were only able to see the citation number and the bibliographic information. By clicking on the hyperlinked citation number, the full-text of that citation would appear and the subject librarians could easily rate the citation on the rubric.

Have full text appear on this page after click to simulate linking from provided document.

How does the acquisition and use of a second language in children affect their general cognitive development?							
Citation Number	References	1 = Below Average Quality		2 = Average Quality		3 = Above Average Quality	
		Accuracy	Authority	Objectivity	Currency	Coverage	Relevancy
01	Caporali, S. A., de Lacerda, C. B. F., & Marques, P. L. (2005). Teaching sign language to the families of the deaf: Focusing on the learning process. <i>PRO-FONO: Revista de Atualizacao Cientifica</i> , 17(1), 89-98.	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
02	Francis, N. (2002). Modular perspectives on bilingualism. <i>International Journal of Bilingual Education and Bilingualism</i> , 5(3), 141-161.	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3

**Accuracy:** reliability, fact checkers/editors, peer review

**Authority:** author's qualifications, reputable publisher

**Objectivity:** minimum bias, extent to which persuasion is the goal

**Currency:** information up to date, date of publication indicated

**Coverage:** depth of coverage

**Relevancy:** related to research topic

## Rubric and Full Text

This screen shows the rubric that we used. It is based on a rubric that has popularly been used to evaluate print resources (Alexander, 1999)

Alexander, J. E. (1999). Web wisdom: How to evaluate and create information quality on the Web.

We asked each subject librarian to assign a score of between 1 and 3 within 6 different categories to each of the citations (1 was below average, 2 was average and 3 was above average).

These six categories were:

Accuracy – which looks at

Authority – specifically the

Objectivity – looking for

Currency – is the information up to date?

How deep is the Coverage

And finally Relevancy – how well does the citation relate to the research question

This resulted in a total possible score of 18 for each citation – we called this a scholarliness score

## Methodology

$$\text{total scholarliness score} = \mu + E_i + L_j + EL_{ij} + \epsilon_{ijkl}$$

where

$\mu$  = Average total score

$E$  = Effect due to exclusivity ( $i = 1, 2, 3$ )

$L$  = Effect due to librarian ( $j = 1, 2, \dots, 7$ )

$EL$  = Interaction between exclusivity and librarian

$\epsilon$  = Error term

We used this statistical model to evaluate the data. Essentially this formula says 2 important things about the way we used the data:

1. We controlled for the differences between the way librarians grade
2. We controlled for the differences in how exclusively the citation was available

This allowed us to pinpoint and measure any differences there may have been between disciplines in our data as well as any differences that can be attributed to the source of the citations

## Results

Google Scholar was 17.6% more scholarly

Citations found only in GS had, on average, a 17.6% higher scholarliness score than citations found only in the DB

## Results

Highest scholarliness score  
when found in both

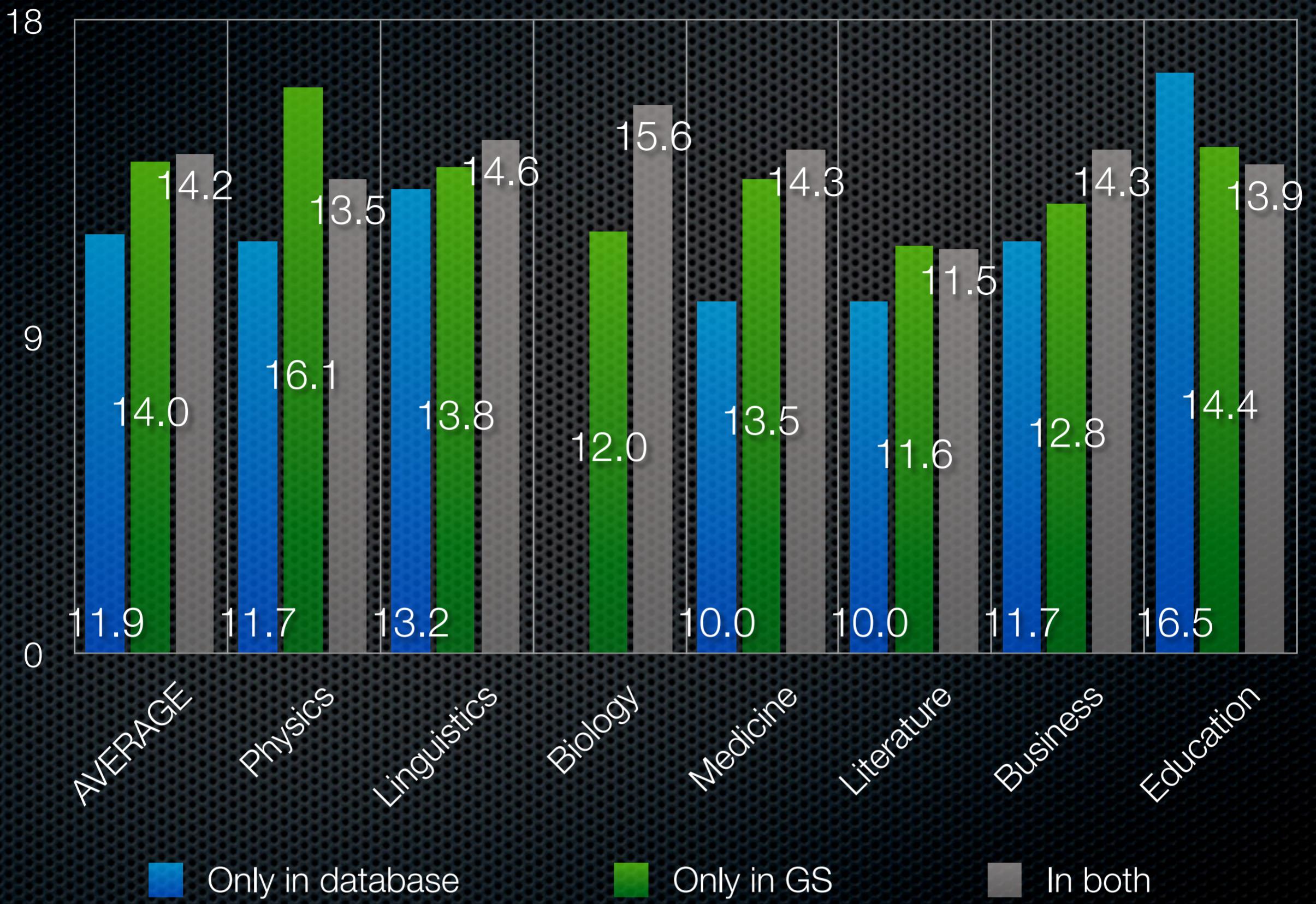
Citations found in both GS and the DB were even higher than citations found only in GS

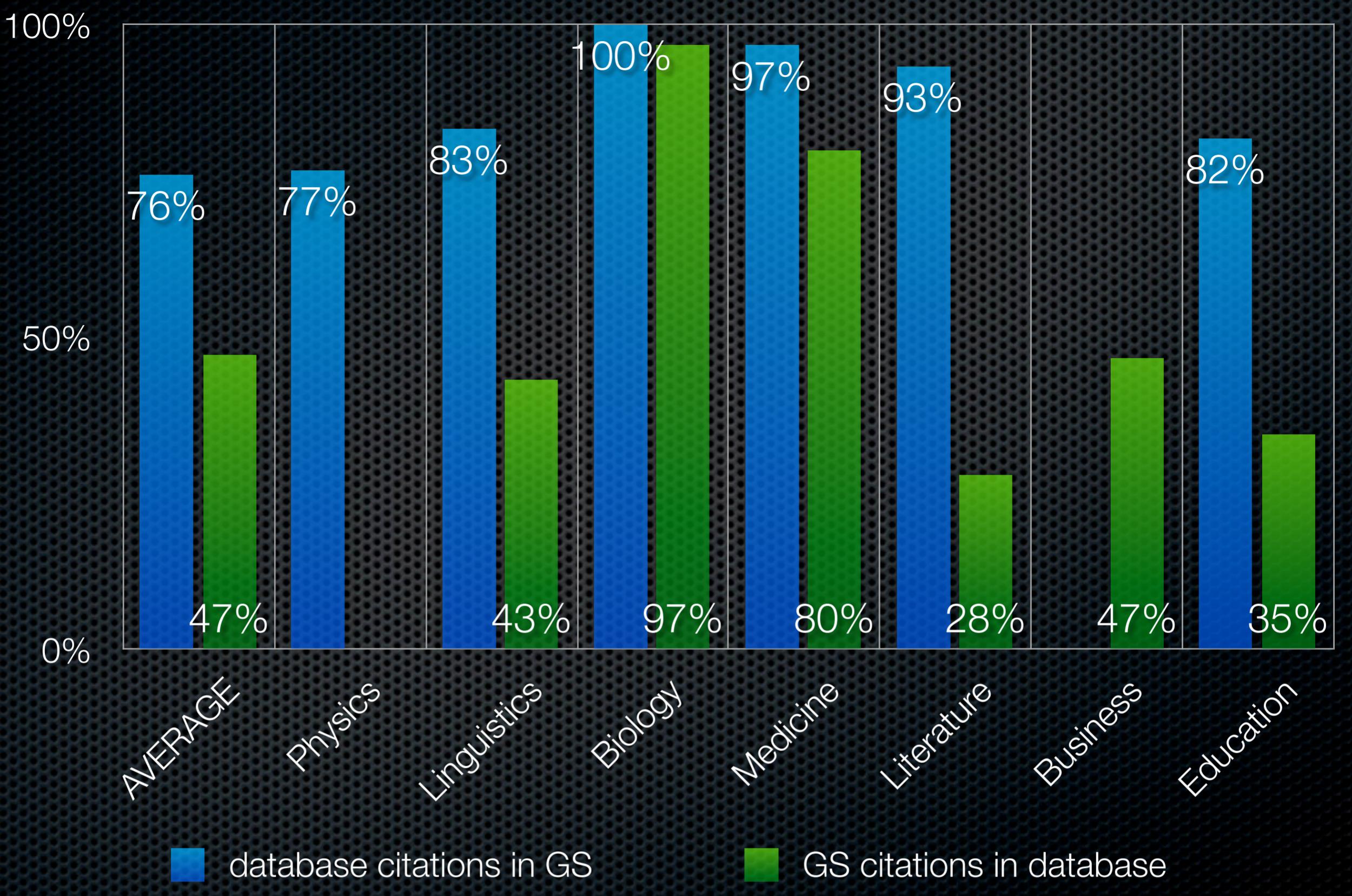
## Results

# No difference between disciplines

We found no statistically significant difference in the scholarliness scores between disciplines (ie, humanities citations in GS are just as scholarly as science citations found in GS)

## Average Scholarliness Score





Future Studies

## Generally applicable results

This study can only be extrapolated statistically to the specific topics and subject specialists used in this study

A more robust statistical methodology would need to be employed to make these results generally applicable

We are encouraged by the results we received and feel that they would probably hold up but cannot say so until another study is done

Future Studies

## Improved rubric

If we had to do it over again, we would have increased the Likert scale on our rubric from 1-3 to 1-7 or 1-10

This would have allowed for a more nuanced statistical analysis and made it easier to spot significant differences, if any, between GS and databases

Future Studies

## Scholarliness calculation

Our scholarliness calculation, ultimately, was based on the subjective opinions of librarians with subject expertise.

There are lots of ways to create a scholarliness score (citation counts, impact factors, etc). Which is best is still debatable

Future Studies

## Comparison to federated searching

Our study compared GS to individual library databases. A more appropriate comparison may be GS to federated search tools.

# Questions?

jared\_howland@byu.edu

tom\_wright@byu.edu

rebecca\_boughan@byu.edu

<http://dspace/byu.edu/handle/1877/634>