



From Citations to Stories: Research into Research via Citation Analysis

By Kerrie Stevens

Director of Library Services, Alphacrucis University College



Introduction

This paper considers two library-led research projects at Alphacrucis University College (AC). As AC works towards becoming a university, research output is a high priority. Universities need to be able to show that a high degree of quality research is being undertaken at the institution. AC has been tracking its research output for years to prepare for university status.

The AC Library has an interest in this aspect of AC's future aspirations in that it is the vehicle by which researchers obtain, seek and request the resources they need to advance their research projects and produce quality outputs. This interest has developed into two related research projects investigating how well the AC library is supporting both research students and staff. As AC works toward becoming a university in Australia, the library can demonstrate support of AC's researchers and their research needs. Every thesis and research publication tells a story – but the footnotes and reference lists tell our story; the story of what libraries have provided, and what researchers still need.



Citation Analysis

Traditional citation analysis (or citation tracking) evaluates research by tracking how often publications are cited by others. It involves tracking and analysing citation patterns to understand the relationship between different works. Usually, publications are given impact scores and other indicators to demonstrate their research value. Highly quoted/referenced citations have a higher ranking and may be seen as more influential/important. However, whilst useful, this approach does not serve library service evaluation.

The method of citation analysis used in these projects is slightly different. It focuses on reference list or bibliography citations and determining whether or not they are held at or are available from the AC library collection. This method helps to determine how well the library is supporting research being undertaken at AC.

By looking at what researchers have used to resource their arguments, libraries can determine if their collections are adequate or if there are areas for improvement to avoid future researchers needing to go elsewhere to resource their research properly. I looked at the titles, age and type of citations being referenced and evaluated the library availability of those titles.

Thesis Project

The first project has been underway for some time. During 2020-2021 work-from-home Covid-19 lockdown periods, I wondered how I could demonstrate the library's ability to support research students when writing their theses. I began to look at citations from thesis reference lists and searching for library availability of these titles.

In late 2024, I became aware of the Excel Alma Lookup Tool²⁵, developed by Thomas Ventimiglia, from Princeton University. We have had several interactions since, and I have been using the new updated version of the App which allows for combined index searching to improve the accuracy of results.

Since AC first started awarding doctoral theses in 2017, there have been 23 PhD, 8 DMin, and 5 MPh theses, totaling over 11,000 citations combined (to 30 June 2025).

²⁵ <https://developers.exlibrisgroup.com/appcenter/excel-alma-lookup/>

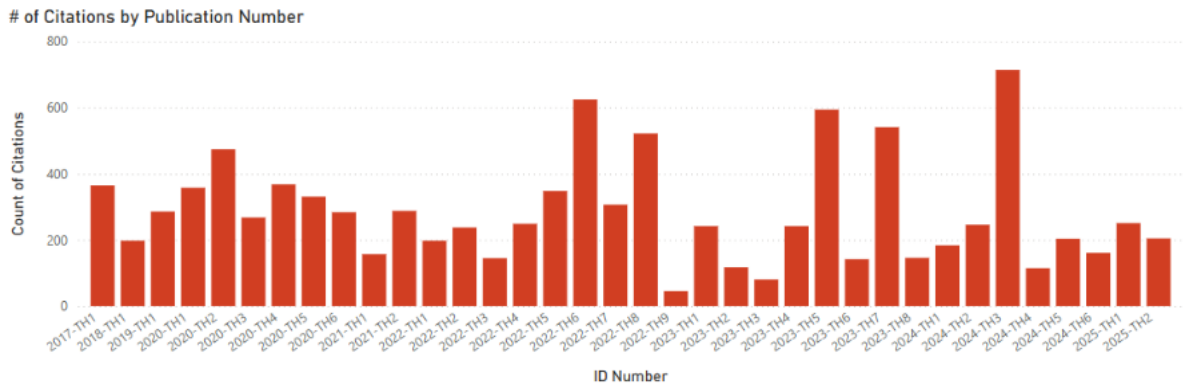
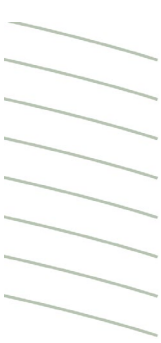


Figure 1: # Citations per thesis 2017-June 2025

Figure 1 shows the wide variation in the number of citations used by each of the theses at AC. Comparison with staff publications will consider only those theses from 2021-2025.

Staff Publications Project

The second project was to consider citations used in staff research publications. Initially, I limited this to publications authored by permanent faculty published in 2025, but I later expanded this to look at all book, journal article and book chapter publications since 2021 as a more comparable basis for discourse with the thesis results over the same time period. Korean language publications were not considered at this time.

Of the staff research publications published between 2021-June 2025, there were 34 articles, 3 books, 11 book chapters and 1 edited book published, using over 12,780 citations combined.

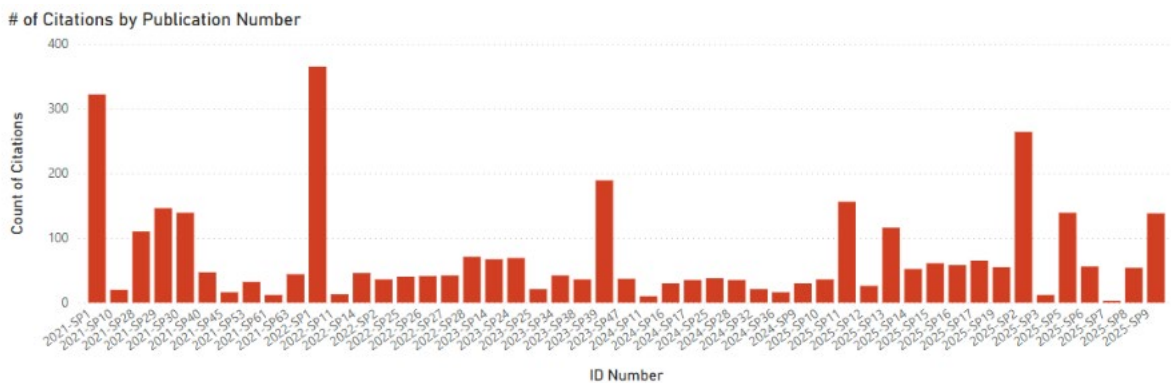


Figure 2: # Citations per staff publication 2021-June 2025



Methodology

Initially, the workload was significant, with over 35 theses from the previous 5-6 years from which to gather data. Now that task is done, ongoing analyses will be undertaken as theses are submitted to the library's research repository and the workload will be easily manageable. However, the data gathered will continue to add to the rich evidence of how the library is resourcing staff and students.

The process is multi-stepped:

- 1 – Extract citations from theses and publications
- 2 – Edit & bullet point citations
- 3 – Run citations through custom app
- 4 – Extract titles from citations in Excel and allocate citation type
- 5 – Run Excel Alma Lookup Tool for library holdings & availability of titles

Whilst there is still some manual work required for the project to progress, this has been reduced significantly from the initial days of this type of citation analysis thanks to a custom-designed app developed by AC's IT Web Applications Support Analyst, Ramon Araujo. His app automatically extracts each citation from a bulleted PDF document into its own row in an Excel spreadsheet allowing for further analysis and data extraction to be done for each citation. The app does speed up the process significantly by automatically extracting the references, citation date, and calculating citation age for each citation. However, it cannot extract the title and author information or determine the citation type – a human is still required for this.

There are also the anomalies brought about by human error, such as errors in the citations, missing or incorrect information, inconsistencies in title presentations, and more, which all contribute to the continuing need for a human eye to look over the citations once in Excel. Also, it seems every thesis and publication uses its own version of referencing style and trying to determine the exact components of a citation is sometimes quite challenging. So, whilst the manual labor was reduced, manual checking of citations, and other data before running the Excel Alma Lookup Tool is still necessary in order to ensure consistent holdings results.

Citations from each publication were divided into the following categories:

Category	Definition
Books	Physical books, ebooks, Kindle books, book chapters, monographs, monograph series
Journals	Articles, e-journals, periodicals, open-access journals, newspapers
Dissertations	Dissertations or theses (published or unpublished)
Web	Online resources, websites, freely available resources on the web, YouTube, Twitter/X posts, TED Talks
Other	Unpublished materials, conference proceedings, Papal documents, research reports (incl PDFs), occasional papers, sound recordings, computer files, sermons, DVDs, CDs, any other format not covered by previous categories

Table 1: Citation Types

Key Findings

The first key finding of these research projects is that both theses and staff publications used very similar percentages of the same types of resources.

Citation Type	Theses	Staff Publications
Books	50%	52%
Journals	35%	38%
Dissertations	3%	1%
Web	8%	3%
Other	4%	6%

Table 2: Citation Types used in Theses and Staff Publications

Of all book, journal and dissertation citations used in theses and staff publications between 2021-June 2025, the library held 43% overall. Web resources are, by definition, available from the internet so were

not included in library holdings. Other items, many of which are also available online, were also not included as they are not resources many libraries typically collect.

Of these citation types held by the library, the breakdown was 49% of them were books, 46% were journals and 3% were dissertations.

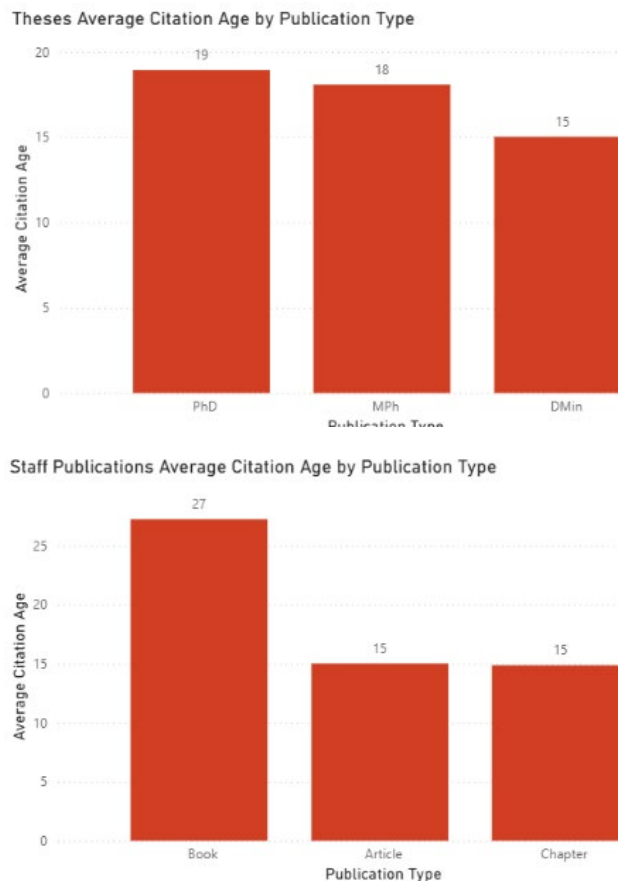


Figure 3: Average Age of Citation Types Used

Whilst both theses and staff publications used similar numbers of citation type resources, the age of those resources differed more significantly. Staff book publications used older references (27 years) whereas theses were fairly similar across the citation age of citations used regardless of the type of thesis being prepared (19-18-15 years). See Figure 3.

Interestingly, staff book publications used much older resources than journal articles or book chapters. This could perhaps be explained through book publications providing the luxury of delving into the history of ideas a little more than a single chapter or a journal article which usually look at more recent ideas.

Average # of Citations for Publication Type

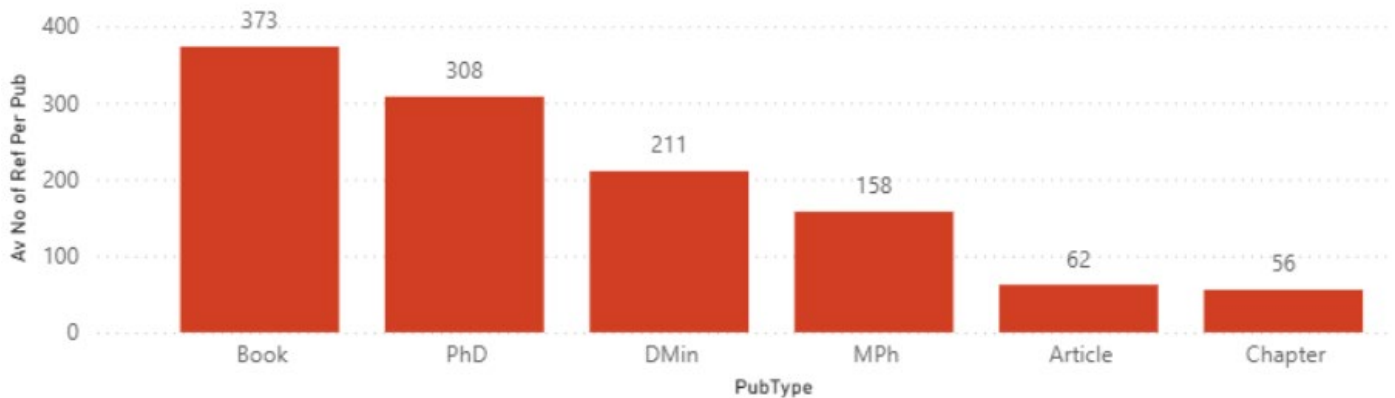


Figure 4: Average Number of Citations per Publication Type 2017-2025

As would be expected, book publications and PhD theses use more citations than other thesis types, journal articles or other publication types due to them being larger projects with the ability to explain and expand on ideas more so than more focused publications.

So, how well does AC Library support research by students and staff? Of the 23,000+ citations used across all publication types, 45% are held by or available from the AC library. There are limited similar studies to be able to accurately compare if this is a good result, but from a personal perspective, if the library can supply almost than half of what researchers need, that is a good start. Now that we know what the situation is, we can work towards improving the rate and move towards a higher percentage of supplying required resources in future years.

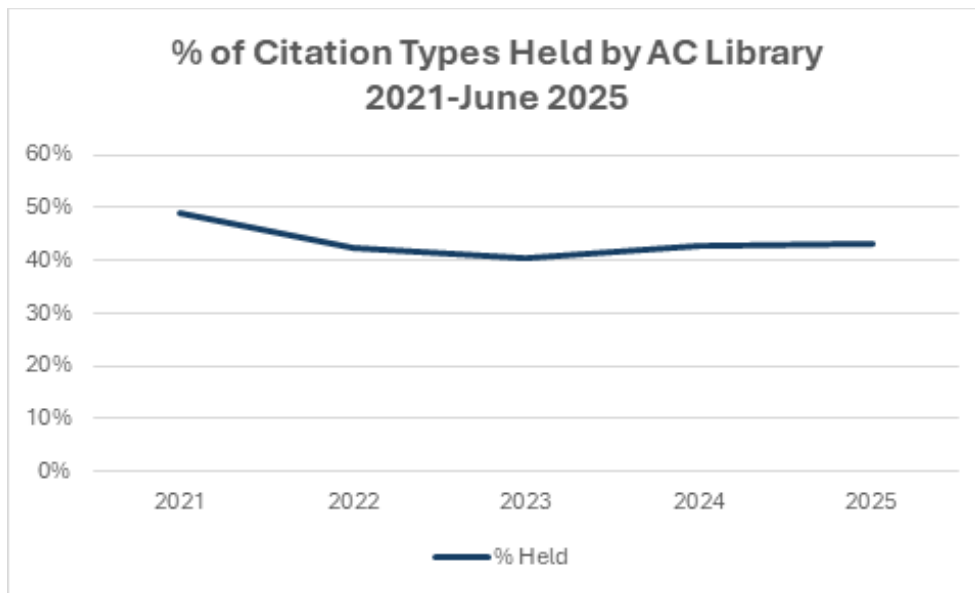
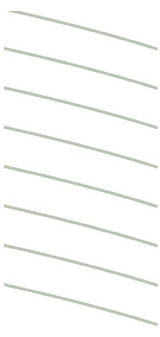


Figure 5: Trend of Citation Availability from AC Library

Looking at the trends of book, journal and dissertation citation types available from AC library by publication year in Figure 5, we can see it has decreased slightly since the initial analysis of 2021 (49% in 2021 to 43% covering Jan-June 2025). I would suggest that 2021 covered the Covid-19 work-from-home period so staff and students would be utilizing AC library’s resources more regularly as it is likely they couldn’t access other libraries as easily. Surprisingly, and perhaps somewhat disappointingly, it seems the ability of AC library to supply researchers resource needs has been declining since 2017²⁶ but thankfully appears to be on an upward trend since a low of 40% of citation types supplied by AC library in 2023. Now that we know this, we can be working towards supplying more of the resources our researchers need when they need them. Whilst this could be seen as discouraging, there are possible reasons for it explained below.

²⁶ This may be due to the limitation of this study looked at library holdings at the time of when the analysis was run, not when the actual publication was being researched. Naturally, the library would have purchased more resources than what would have been available at the time of earlier research publications. Future analysis will be closer to real-time with regards to library supply of citations used as the lapse of time between the publication and checking of library holdings will be greatly reduced. See Limitations & Assumptions section for further discussion.



Limitations & Assumptions

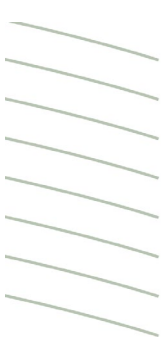
As with any research project, there are numerous limitations and assumptions that must be considered and acknowledged.

First, the library holdings represent what is in or available from the library when the Excel Alma Lookup Tool was run, not what was available when the thesis or research publication was being written. This was more obvious when the earlier theses were first considered (eg: 68% of citations used in 2017 theses were available from the AC library in 2021, 4+ years after the theses were written), but now that I am up to date with citation harvesting, any new theses or publications being analysed can be confident of more accurate library holdings at the time of writing because there is not such a great time difference between the publication and the checking of library holdings.

Second, the accuracy of the Excel Alma Lookup Tool was initially questionable because it was only able to search for title and date information. This was fine for most book titles if they had unique titles, but for a theological library with dozens of commentaries named by the Bible book name under consideration, this was problematic. For example, a commentary called 'Mark' published in 2023: we have 3 books by different authors, but the first iteration of the Lookup Tool could not differentiate between them so I could not confidently determine if the exact title being referenced was available from the library collection. With access to the updated version of the Excel Alma Lookup Tool, this was resolved by being able to include the author/editor names in the search as well.

Third, the analysis is retrospective, looking at library holdings for research already published or completed, sometimes several years ago. Whilst the analysis doesn't benefit those projects, it does help future research projects to be well-supported by library resources going forward. Cumulatively, each year of research publications benefits from the previous year's citation analysis as library holdings are developed to better support both staff and student research interests at AC.

Fourth, simply determining the availability of a title in the library collection does not necessarily mean the researcher accessed it from the library. As some holdings may not have been available, especially for earlier theses, it was simply assumed that access to cited titles was via the library regardless of when the library may have acquired access to a particular title. Also, regarding journal holdings, the Lookup Tool was not used to determine the range of holdings held, so, whilst we may hold a particular journal title, we may not necessarily hold the required volume and issue number referenced in a thesis or staff publication. For the purposes of this citation analysis, if we held the journal title, it was considered available from the library even if we may not have held the actual article used.



Fifth, many of our staff researchers hold positions at other institutions as well or they have co-authored a paper with someone from another institution. This would mean they would likely have access to that other institution's library resources as well. For the purposes of this analysis, holdings or availability of other libraries was not considered. This would likely bring down the overall percentage of resources available from AC library, particularly for some staff publications covering subject areas which are not collection priorities of the AC library. Papers co-authored by multiple AC staff members and/or research students were considered only once.

Benefits

This project has and will continue to benefit the AC library and students/staff in a number of ways:

- Future students will have access to an ever-growing research level collection based on previous researcher requests and requirements.
- Gaining a better understanding of the types of resources researchers are using helps identify gaps in the library collection and inform future purchasing decisions.
- All students benefit from the RapidILL service that was introduced in early 2025 to provide free access to journal articles and book chapters for all staff and students. Previously, only HDR students and staff were provided with free interlibrary loans and document delivery.
- The library has raised its profile within AC, being invited to join the Research Committee. Collection development decisions can be made with real evidence of need.

For smaller or theological libraries, the benefits of undertaking a similar citation analysis can be three-fold:

1. Gain a better understanding if your library has what your researchers are using in their research projects and publications, and then take on the possibility to do something about it if you don't;
2. Take advantage of the amazing and free data source available right on your shelves – those shelves of theses contain a wealth of information about your collection; you just need to extract it... and
3. Citation analysis as described here is a low-cost, high-impact method to demonstrate library support for students and staff. It does take a significant amount of time initially, but once set up and up-to-date, ongoing analysis is almost cost neutral.



Next Steps

A small bench-marking project was undertaken with Avondale University Library in early 2025 looking at their theses from 2017-2024 and their library holdings. On par, AC library was able to demonstrate equivalent provision of resources used in theses across the same time frame as another institution that has already achieved university status.

I would be interested in undertaking further similar projects with other libraries who offer doctoral degrees in the future. The Excel Alma Lookup Tool can also extract holdings data from OCLC WMS catalogues so I would be interested in exploring this with others in the future for the purposes of comparison and information sharing.

Conclusion

Through the analysis of citations referenced in both theses and staff research publications, this project has determined that the AC library is supporting AC researcher's needs. Identifying the current rate of supply of citations means we have a starting point to use as impetus to improve library supply of used citations. This will benefit the institution's goal to become a university in coming years. It has also identified areas where collection gaps may exist, and further support may be required. In conclusion, I hope this presentation has opened your eyes to the rich data you have available at your fingertips that just needs a little effort to produce great benefits. By analysing what our students and staff are using, we can show decision makers that we are, or are not, adequately supporting our institution constituents' research needs. And that may be the evidence you need to request funding for another database, further resources or something else your researchers need.