

**INTERPRETING iTHENTICATE™ REPORTS: A GUIDE FOR RESEARCHERS**

Table of Contents

[Introduction 1](#_Toc475976099)

[Percentages 2](#_Toc475976100)

[Interpreting percentages 2](#_Toc475976101)

[Low percentages 2](#_Toc475976102)

[High percentages 2](#_Toc475976103)

[Document sections 2](#_Toc475976104)

[Percentages and filters 3](#_Toc475976105)

[The Similarity Report 4](#_Toc475976106)

[Viewing modes 4](#_Toc475976107)

[Example and notes 4](#_Toc475976108)

[Explanation of key points and features 5](#_Toc475976109)

[Key questions 6](#_Toc475976110)

[Is it plagiarism? 7](#_Toc475976111)

[Typographical errors 7](#_Toc475976112)

[Poor paraphrasing 7](#_Toc475976113)

[Standard phrases 7](#_Toc475976114)

[But I haven’t read that! 7](#_Toc475976115)

[Self-citation and self-plagiarism 9](#_Toc475976116)

[Key questions 9](#_Toc475976117)

[Genres of writing 10](#_Toc475976118)

[PhD or Masters Thesis 10](#_Toc475976119)

[Journal articles 10](#_Toc475976120)

[Books 11](#_Toc475976121)

[Blog posts 11](#_Toc475976122)

[Not Just Plagiarism 12](#_Toc475976123)

[Reference checking 12](#_Toc475976124)

[Copyright and permissions 12](#_Toc475976125)

# Introduction

iThenticate software compares uploaded documents with published scholarship and webpages. It identifies text which matches material that has already been published. An overall percentage score is generated, along with a detailed report. Researchers need to pay close attention to the reports it generates to determine whether matching text is a problem which needs to be addressed. This guide has been prepared by the Research Education and Development team to help researchers make the most of iThenticate.

Matching text is not necessarily plagiarism: it may be correctly cited, or use standard or general phrases for example. iThenticate can help researchers avoid inadvertent plagiarism of others’ or their own previously published work. It can also be of use in other aspects of the writing and publication process, such as addressing copyright regulations.

iThenticate does **not** add any documents you scan to its database of sources (unlike Turnitin). This means that your intellectual property is protected when you use it.

This guide helps researchers understand iThenticate Similarity Reports and how they can be used to improve scholarly publications. It assumes that the user is familiar with the software. If users are unsure about any terms or how to perform particular tasks an introductory guide to the software is available here.

Although iThenticate is a powerful tool, using it does **not** guarantee that writing does not contain plagiarised text. It does not check against, for example, Google Books or material that has only been published in hard copy. It is essential that using iThenticate is only part of a researcher’s strategy to ensure academic and research integrity.

## Copyright @ La Trobe University

Copyright in the information in this manual is vested in La Trobe University pursuant to the Australian Copyright Act 1968. Unless otherwise stated, no part may be reproduced by any process, unless permitted by the Australian Copyright Act 1968 or used for any commercial purposes without the written permission of La Trobe University.

## Disclaimer

While every effort has been made by La Trobe University to ensure that the information in this manual is up to date and accurate, La Trobe University does not give any guarantees, undertakings or warranties in relation to the accuracy, completeness and up to date status of the information.

La Trobe University will not be liable for any loss or damage suffered by any person arising out of the reliance of any information in this manual.

La Trobe University accepts no responsibility or liability for the content available at any of the web sites referred to in this manual.

## Trademark

iThenticate is a trademark of iParadigms, LLC.

# Percentages

When a document is scanned iThenticate calculates a percentage score of the number of matching words from all identified sources compared to the total number of words in the document. More information about how the score is calculated can be found [here](http://www.ithenticate.com/plagiarism-detection-blog/bid/63534/CrossCheck-Plagiarism-Screening-Understanding-the-Similarity-Score#.WJuaXYVOK71).



This is the first piece of information researchers are given about their results. Without careful interpretation, however, the percentage score and report have no particular significance.

## Interpreting percentages

There are no standard or universal ideal maximum or minimum percentages to aim for. Even a document with a low percentage could contain plagiarised text, while a document with a high percentage may have no problems.

### Low percentages

A document with a low percentage score (less than 20%) should still be carefully checked to ensure that citations are correct and that matching text is attributed correctly. Use the Similarity Report to do this.

If the score is low researchers may wish to consider whether the document engages closely enough with existing literature in the field.

### High percentages

 A medium to high percentage score (above 20%) requires more consideration. A higher percentage score can result from either:

1. significant quotation from a single source; or
2. a combination of many shorter quotes from multiple sources.

Both of these may be acceptable according to context and usage. An English article about a single poem or novel would likely quote extensively from the primary text, while a systematic review in medicine or the sciences might quote briefly from many sources.

The more quotations are used the more likely it is that inadvertent plagiarism will occur through errors in note-taking, typing, etc.

Even if all matching text is quoted correctly, a high percentage score may indicate that the document does not meet academic standards for originality of ideas, argument, or new discoveries. See the Self-citation and Self-plagiarism and Genres of Writing sections of this guide for more information about the levels of originality expected in different kinds of publications.

### Document sections

The different sections of a document may have varied levels of text matching. A high percentage of correctly cited matching text is more likely to be acceptable in literature reviews, methodology, and materials sections than in discussion or conclusions sections.

It is recommended that filters are **not** set to “Exclude Abstract” or “Exclude Methods and Materials”. It is important that all matching text is identified and reviewed.

## Percentages and filters

It is important to set filters carefully to ensure percentage results are not misleading and to avoid having too much text highlighted for a report to be read effectively. The two images below show results for the same document. The first has the filters set to “Exclude Quotes” and “Exclude Bibliography,” while the second does not.





Filters can also be set to exclude sources and/or matched that are less than a selected size. This is not generally recommended as it is likely to prevent iThenticate from identifying errors including poor paraphrasing and [mosaic plagiarism](http://isites.harvard.edu/icb/icb.do?keyword=k70847&pageid=icb.page342054).

See Topic 5 – Report Filters and Phrase Exclusions of the Introduction to iThenticate guide for information on how to set filters.

Note that when using iThenticate to assist in checking references and/or copyright compliance the “Exclude Quotes” filter box must **not** be ticked.

# The Similarity Report

Access reports by double-clicking on the percentage score of a document. These can be viewed in either Document View or Text Only. Toggle to Text Only by clicking the button in the bottom right corner of the window. Toggle back to Document View by clicking the ‘Document View’ button at the top left of the Text Only view page.

## Viewing modes

Text Only View has some modes which are not available in Document View, but Document View is easier to compare with a file or printed copy. Document View in Match Overview mode is recommended unless an alternative mdoe is identified for particular tasks.

The two modes available in Document View are:

1. Match Overview, which gives a list of ‘best’ sources for matching text according to the iThenticate algorithm. This option has the least cluttered view.
2. All Sources, which gives a list of all sources with matching text. This may include the same webpages cached on several different days, and other unnecessary or confusing matches. This option is recommended for checking citations are correct.

In Text Only view:

1. Content Tracking mode highlights all matching text from a single source (selected from a list of all sources) throughout the document, and provides an accurate count of all matching words. This option is recommended for checking copyright compliance.
2. Summary Report provides the list of matched ‘best’ sources at the top of the image, with the document and corresponding highlighted text displayed underneath it. This option can be useful for comparing against a bibliography or reference list to ensure completeness.

Further information about using viewing modes is available [here](http://www.ithenticate.com/products/dv-notes) and in Topic 4 of An Introduction to iThenticate.

## Example and notes



G

F

**E**

**C**

**D**

**A**

**B**

This example is in Document View, Match Overview mode. Each section of matching text is highlighted. The colour and number match the relevant source in the Match Overview column on the right of the image.

The overall percentage for this short section of text (36%) could have been reduced to 19% by setting the filters to “Exclude Quotes” and “Exclude Bibliography” as in the example on page 4 of this guide.

## Explanation of key points and features

1. A correctly paraphrased sentence cited at endnote 1. The text does not match so it was not highlighted by iThenticate. **Action:** none needed.
2. Two sections of problem text are highlighted as occurring in source 1 in the Match Overview column.

No source is given for the first instance, which is almost a complete sentence and appears to have been ‘cut-and-pasted’. This may be considered plagiarism at worst, or sloppy work at best. **Action needed:** insert quotation marks and a citation.

The second section from source 1 (in the last line of the paragraph) gives a citation. It is, however, a very poor example of paraphrasing as the last part of sentence has a substantial number of words in common with the source. **Action needed:** edit the section of text which has highlights to improve the paraphrasing.

1. Some text within a quote – which is cited correctly via footnote 2 – is highlighted as matching words from the *Handbook of Academic Integrity* (source 4 in the Match Overview column). This is not plagiarism on the part of the author of the piece of writing above, or the authors of the source that is quoted. This is an example of a standard or generic phrase. The highlighted words “students to avoid plagiarism” does not express an original idea or previously unknown fact. Nor is it distinctive. **Action:** None needed. Changing the filters to “Exclude Quotes” would remove all the unnecessarily highlighted text in this sentence. Note that standard phrases will not always be embedded within a quote.
2. Source 3, highlighted in purple, matches text in the endnotes. This source is another piece of writing in the field that cites the same sources using the same referencing style. **Action:** Change filters to “Exclude Bibliography”.
3. This figure is **not** always the total matching words to a source when in Match Overview mode. Go to Text-Only Report and select Content Tracker mode to see the **correct** total words matching each source. See also the section on Copyright and Permissions below.
4. Click this button to set filters in Document View. In Text Only view they appear on the top right of the screen. Filters may adjusted at any stage. They will need to be set differently to enable checking for different kinds of issues.
5. Click here to create a PDF of the report for reviewing offline or printing. Note that the Match Overview will appear in the pages **after** the document with highlighted text rather than beside it as in the online version. If All Sources mode is in use only those sources which have been selected (by clicking on them in the right-hand pane) will be printed. Printing/ converting to PDF in Match Overview mode is recommended.

## Key questions

Researchers should review each section of highlighted text. Below are some guiding questions that should be considered.

1. Is all matching text enclosed in quotation marks and correctly cited?
2. Are all quotations attributed to the correct source?
3. Are clusters of quotations in appropriate sections of the document?
4. Has any single source been quoted to an extent that copyright permission should be sought?

# Is it plagiarism?

There are many different kinds of plagiarism. It does not simply consist of cutting and pasting large chunks of another author’s text and claiming them as one’s own. [This guide](http://turnitin.com/assets/en_us/media/plagiarism_spectrum.php) from the makers of iThenticate outlines some of the most common. Harvard University also has a [guide](http://isites.harvard.edu/icb/icb.do?keyword=k70847&pageid=icb.page342054) that includes examples of different kinds of plagiarism and corrections to them.

This section gives some common examples of errors that may be identified by iThenticate.

## Typographical errors

iThenticate can help identify mistakes made when quoting sources (such as getting a word wrong in a log quote), and in typing (such as omission of opening or closing quotation marks).

Make sure filters are set to “Include quotes” to check for typographical and transcription errors in quoted text.

## Poor paraphrasing

Sometimes errors occur when researchers paraphrase poorly and include words or phrases from their source. iThenticate may highlight these mistakes, as in point B of the Similarity Report example above. Even if a citation is included the text should be edited to remove the repeated words and create a correct paraphrase.

## Standard phrases

All disciplines have particular phrases that are very commonly used and may not require attribution. There are only so many ways in which a particular methodology or material may be clearly explained, for example.

The phrase “perform a close reading”, which describes a Humanities methodology, appears in 407 papers listed on Google Scholar, but is not in quote marks in any of them. This does not mean that 406 of them have plagiarised the first published. The key question in this type of example is whether the phrase is a standard one or is original to a specific source. This must be considered on a case-by-case basis by the researcher.

Any matching text should be carefully considered and rephrased if possible to avoid any appearance of plagiarism or lazy writing.

Phrases found in the University of Manchester [Academic Phrase Bank](http://www.phrasebank.manchester.ac.uk/) are highly likely to be highlighted as matching text. Using them does not constitute plagiarism.

## But I haven’t read that!

iThenticate may identify matching text from a source or sources which is unfamiliar.

If you do not recognise the source in the Match Overview pane, switch to All Sources view by toggling the column graph icons. Match Overview displays only the ‘best’ match for each piece of text as determined by the iThenticate algorithm; All Sources displays a complete list of identified sources which have matching text.



If no items in the All Sources list are recognised, records (such as an Endnote database) should be checked closely. A citation manager may repeat errors that occur in initial entry of sources, and metadata (used in automatic entry of downloaded files) may be incorrect.

The text may need to be revised either by editing or by engaging with the identified source and referencing correctly so that there is no appearance of plagiarism. A publication that has significant levels of matching phrases and is from the same field may be relevant even if it is previously unknown. Researchers should remember that engaging with existing scholarship involves representing the current state of the field accurately as well as giving correct citations.

See also the section on Reference checking on page 12 of this guide if the matching text is a quote with a citation attached that does not appear in the All Sources list.

# Self-citation and self-plagiarism

Researchers commonly write multiple publications on a topic over a period of time. It can be easy to self-plagiarise inadvertently by using the same text.

Different genres of academic writing have different conventions about the level of originality required. Journal editors expect a high degree of originality and are unlikely to look favourably on a submission that reuses substantial amounts of text the author has published before even if it is cited. In contrast, a monograph publisher may accept a chapter substantially based on an article that is already published.

iThenticate can help researchers avoid self-plagiarism and check that the level of originality in a piece of writing is appropriate to the planned publication venue by identifying text which matches their previous publication(s) word-for-word.

To make scanning a document which is intentionally based on a previous publication (e.g. thesis chapter based on an article) for other issues easier, go to Match Breakdown for the relevant source by clicking the arrow which appears when the mouse pointer is hovered on the right of the source in Match Overview.

##

Next click the “Exclude Sources” button at the bottom of the right hand viewing pane, then select the relevant source(s) by checking the boxes next to them. The key questions below should be considered for each source *before* it is excluded.

## Key questions

When interpreting a Similarity Report there are a number of critical questions to consider in relation to self-citation and self-plagiarism for each section of text that matches a researcher’s previous publications (sole or co-authored).

1. Can this section of text be edited sufficiently to avoid matching? If yes, a reference should be provided to indicate where the paraphrased ideas come from.
2. Has any re-used text been cited correctly?
3. Is it appropriate to re-use text from the original genre of publication in a new one? (See the Genres of Writing section below)
4. Who owns the copyright on the original text? Has permission to use it been given? Copyright and Permissions section below)
5. Is the previously published work sole or co-authored? Have all co-authors agreed to re-use? (See the La Trobe [Authorship of Research Outputs Policy](https://policies.latrobe.edu.au/document/view.php?id=105%20?))

More general information is available on the Self-citation and self-plagiarism webpage.

# Genres of writing

Different genres of writing in different disciplines have varied requirements. Researchers should consider carefully the type of document they are working on when interpreting the iThenticate Similarity Report.

Researchers should discuss their draft(s) with a supervisor or mentor in their discipline if they are not confident that the requirements of the genre of writing and discipline have been met.

This section outlines some issues for common types of academic writing.

## PhD or Masters Thesis

Many candidates publish during their candidature and include either sections or entire publications in their thesis. It is important that all sections of a thesis which have been previously published are correctly acknowledged and the permission of co-authors obtained where relevant. Permission from the copyright holder may also be necessary.

If candidates are quoting sections of their own publication (single or co-authored) it should be referenced as the work of another scholar would be.

Section B of the [Schedule for Presentation of Theses for Higher Degree by Research](http://www.latrobe.edu.au/__data/assets/pdf_file/0005/670541/Schedule-for-Presentation-of-Theses-for-Higher-Degrees-by-Research.pdf) gives details about how previous publications should be incorporated into a thesis. The three key points to remember are that:

1. Full citations must be given in the body of the thesis (.e.g preface or each relevant chapter) AND in the bibliography.
2. Candidates must have permission from the publisher and follow any requirements, including regarding publication of the thesis in the University repository.
3. If a publication was jointly authored the candidate must have taken a leading and significant role, and must include written confirmation of their contribution from co-authors.

See also the Self-citation and self-plagiarism webpage.

It is crucial that candidates abide by copyright regulations when citing their own previously published work. Permission in writing from the publisher may need to be obtained. Many publishers have instructions and/or a form to complete in the ‘Permissions’ or ‘Authors’ sections of their websites.

All use of direct text must comply with the concept of fair use. See the Copyright and Permissions section of this guide for more details.

## Journal articles

Journal editors expect to receive submissions that contain original research and are unlikely to publish work that has already appeared elsewhere. It is common practice for editors to scan submissions with iThenticate or similar software. By using iThenticate on the final version of an article researchers may avoid problems and embarrassment, as well as save time in revisions.

Research based on Masters or PhD theses may be an exception to this general rule, but some journals either do not accept thesis work or only do so if it has been revised extensively. The style of writing suitable to a thesis may not suit some journals, so a high percentage score with a researcher’s own thesis in iThenticate may indicate that substantial re-drafting is needed.

See the Instructions to Authors section of a journal webpage or contact the editor in advance of submission to check for specific policies on reusing text from a thesis or other previous publication. If no information is provided researchers may also contact the editorial team too clarify.

## Books

If an entire book, or sections of one, are based on previously published work such as a PhD thesis or journal articles iThenticate will generate a high percentage score for the document. Whether such a book is acceptable varies significantly across the disciplines, so this will need to be considered according to research context, e.g. discipline. It is vital that researchers are open with any prospective publisher about the status of their work.

Questions of copyright may arise if researchers are substantially re-using their own work. iThenticate can identify exactly which sections of text are reproduced verbatim. See the Copyright and Permissions section below.

## Blog posts

Blog posts are a relatively new form of academic publication compared to journal articles and books. They may be highly original as researchers use them to ‘road test’ ideas before formal publication, or based on existing publications as they can provide an accessible public forum to report specialist findings.

If a post is giving an account of a piece of published research iThenticate’s text-match function can identify text which:

1. Should be rephrased to avoid infringing copyright;
2. Requires a citation and quotation marks;
3. Needs editing to make it more accessible to a non-specialist audience/ the general public.

# Not Just Plagiarism

## Reference checking

Even experienced and careful researchers can make errors when taking notes and referencing, while citation managers can result in repeated errors if a mistake is made in manual entry or metadata. Giving the incorrect source may not be plagiarism, but detracts from the credibility of an argument and creates the appearance of careless research practices.

iThenticate reports can help identify problems of this kind prior to submission for publication.

Ensure that the “Exclude Quotes” filter is **not** set, then check that the citation given in the text is the same one located by iThenticate. If it is not first check the full list of possible sources in All Sources mode. If it does not appear there go back to the original source that is cited in the document to confirm the reference. If it is correct do **not** change it to the one identified by iThenticate as some sources may have been published more than once, for example a journal article included in a collection of essays etc. Researchers should always cite the version that they have used themselves.

Generating a Summary Report will provide a list of the sources of matching text found by iThenticate. This may be compared with a bibliography or reference list as a preliminary check.

## Copyright and permissions

Written permission to quote from copyright work may be required if more than a given number of words are used. The La Trobe library has a [Copyright Hub](http://www.latrobe.edu.au/library/copyright-hub) with a range of helpful resources if you need more information. iThenticate can help check whether written permission should be sought.

To check the number of words from each source go to Text-Only Report view and choose Content Tracking from the drop-down box in the top left corner. The total number of words from each source will be displayed in the right-hand pane. Note that in the example below the Similarity Report algorithm masks some matching words within matches from other sources. Match Overview and All Sources modes in Document Viewer do **not** show accurate totals.



If this exceeds the limit set by the publisher permission should be sought. The published of **either** the original source **or** to which a document will be submitted may have rules with which researchers need to comply.

**Note:** iThenticate may not include an entire source (e.g. a novel) in its database. Always scan the document carefully yourself to ensure that all quoted text has been identified and counted.