# PDF Series. Session 1. What is Text-Readability?

# Video Transcript

## Slide 1 - Introduction

Thanks everyone for joining and this is the first session in our series on PDFs. I'll explain a bit to you in a minute about why we’re doing a series on PDFs. There's a huge context in the university for accessibility around the use of these. They're a really popular type of content in our courses. So it's one thing we really want to show you, some of the little tricks you can do to keep them accessible.

As you would have seen on the sign-up, this is 15 minutes. The intention is that we're teaching you one little bit. For a start. [Admits participant from the lobby].

The idea with this series is that it progresses and gradually gets a bit more advanced, so the next session is about a few things that might trip up a PDF with accessibility. The one after that is starting in your Word document and really just breaking it down to set up that initial Word document so that it’s accessible. The final session would be for users of Acrobat Pro DC. So the licensed version. If you want to remediate things, in that, our final session would be that.

But for today, we're just looking at text-readability, so this mainly pertains to scanned documents, so it it's really all about scanned documents or anything that isn't text-readable as they can trip you up for accessibility.

## Slide 2 (1:34) – Objectives of the Session

So what you will be getting out of this session is an understanding of text-readability and why it's beneficial to learners. And how to ensure your documents are text-readable and how to rectify issues with text-readability when you come across it.

Slide 3 (1:50) – Context of This Session  
So the context of this session came about because we were running a pilot over the summer, this summer just gone, of this software called Ally that plugs into canvas. Now I've a screenshot of it here, how it shows in the Canvas menu. Don't go looking for it right now. It won't be there because we're enabling this after Christmas. We're just doing a bit more piloting with it first, but the example on the right is an example of a report that it shows in a course and you can see that there's 496 issues with content in this course.

What Ally does is it scans your modules in Canvas across all the content and it will tell you what you need to fix to improve the accessibility. So it's going across all your Word documents and PDFs and images and text and all of it. And you can see PDF is right up the top here. There's over 200 issues in this one module on Canvas with PDF documents.

And that's this is an example from one course, but most courses are the same. The PDFs are generally the biggest bugbear for accessibility.

Slide 4 (2:48) – Context of this Session  
This is a list from Ally of issues that show up with documents. So we'll be talking about a lot of these throughout the series. Tagging will be next week, and headings and images, table headers that kind of thing. We will be looking at these throughout the series and you can see there's hundreds of things show up.

The one up the very top here is ‘the document is scanned but not OCR’d’ and that's what we're looking at today. This idea of OCR documents and that you get more familiar with this term and what it means and why it's important.

Slide 5 (3:20) – What is Text-Readability in a PDF?  
So what is text-readability? It’s the name of the session and it's the one little point we're making today. So it means Optical Character Recognition. If a PDF has Optical Character Recognition then it's readable. So OCR is the acronym you’ll see for this. So it basically means the characters in the PDF can be recognised. So when you scan a document it usually comes out as an image of text. So you can see here in the little icon, it represents what an OCR will do, it will make the text-readable.

It's the electronic conversion of images of typed, handwritten, or printed text into machine-encoded text. Machine-encoded means that the PDF knows that it's text, as opposed to seeing it as an image of text. So you can check for this by highlighting the text with your cursor. And I've just a little note to myself here to demo that.

### Demo – Recognising Text-Readability

This is an example of a PDF that has text-readability in it, so I can put my cursor in the way I can in a Word document and page through it and I can grab words. So that's literally it. That's how you test for text-readability. You can highlight the words in the document and that means that it's text-readable.

And this scan here is an example. Now it looks fine to the naked eye. It's quite clear you can read that, but when I go to grab the words, I just get these snap boxes. If I tried to go over a specific word here, it's not. It's treating it as an image basically. So that's a non-readable PDF and scans are particularly dangerous when it comes to this. They're usually the main culprit.

Slide 6 (5:13) – Why do We Need Text-Readability?  
And so why? Why is this important? Why doesn't matter if your text is readable or not?

And so text-readability. It ensures that the text in PDF is searchable and readable by a screen reader. So for those who are blind or have a vision impairment and they're using a screen reader software. You might have heard of Jaws or NVDA. I'll show you a demo in a minute of NVDA. That's a free screen reading software.

A dyslexia support software such as Claro Read Pro or Read & Write. So students with dyslexia might be reading their content on Canvas with one of these. And if you have a document up on that or they might be just going through the downloaded content. A read aloud software such as Immersive Reader. So this comes freely available in Word.

And lots of students, even whether they're registered with the DSS or not, might be using Immersive Reader to have the text read aloud to them, or speak selection in iOS, or a braille device.

Slide 7 – (6:05) Video  
And so this example here, I made a short video (I didn't want to demo the screen reader itself as it's quite clunky to open up) so I've made a short video of what will happen.

This is the document I showed you a minute ago that isn't text-readable, and this is how a screen reader will try to read it. [Presenter re-shares screen to enable computer audio].

And if you listen carefully, it might not be that clear. You'll hear what happens when the screen reader tries to read this document.

**Screenreader**  
“Empty document. This document appears to be empty. It may be a scanned image that needs OCR, or it may be a malformed document.”

You can see that the screen reader knows it’s a document. It's actually kind of guessing itself that it hasn't been OCR’d. It's kind of making a guess. “There could be something here, but I'm not getting it.” So when you make your document text-readable, you're basically telling that screen reader “yes, you've text here and you can read it out.”

Slide 8 (7:18) – Fixing a Scanned Document that was not OCR’d  
So what you do with your scanned documents that don't have the screen readability on them? So there is a way to to fix this.

* **Acrobat Pro DC**, if you're using the licensed version, it has a scanning OCR function.

### Demonstration

For users of the licensed version, there's a little ‘Scan & OCR’ button here. If you don't see it, you'll find it under ‘More Tools’.

* **Google Docs** as well. If you upload a document to Google Docs and convert it to a Google doc, that's a form of OCRing when you can pull out the text.
* **SensusAccess**: but the main tool (and our interns found this over the seven weeks of the project). They were remediating content in modules in the College of Business and Law, and they used this service called SensusAccess that we have in UCC. They found they were using it a couple of times a day.

And they also found that PDFs were the biggest issue. As I mentioned already, and that's why we're running this series on PDFs. So it's a self-service alternate media solution for educational institutions. It allows students, faculty, staff and alumni to automatically convert documents into a range of alternate media formats. So for example, PDF to word, PDF-image to accessible PDF, PDF to Word, MP3 and many more.

Demonstration – SensusAccess

* Go to Libguides.ucc.ie
* Select the ‘Supporting Learning’ tab
* Select SensusAccess
* Select the ‘Convert Files Now’ button
* Source = File
  + **Step 1**. **Upload your Document**
    - Choose your File
    - Press Upload
  + **Step 2**. **Select Output Format**
    - For this demo, we’re using ‘Accessibility Conversion’
  + **Step 3 - Specify accessibility conversion options**
    - Under ‘Target Format’ we can choose ‘Microsoft Word’ or ‘PDF.Text-Over-Image’
  + **Step 4.** **Enter email address and submit request**
    - Enter your UCC email address. This will take a couple of minutes to come through.

Slide 9 (11:04) - Thanks for Listening  
And so that's it, as I said, it's one simple trick that we wanted to show you and the rest of the detail will kind of progress throughout the series.

I'm around for questions now, this is our e-mail address and our website is available here. I'd also like to say thanks to Michelle O'Halloran on the call from the DSS who introduced me to all of this and I used some of your resources from today so thanks very much, Michelle.

## Ally Pilot

And also, if you'd like to join our pilot, so ally, the tool that I showed you at the beginning here, we're doing a second pilot of this.

[Ally Phase 2 Pilot Webpage](https://www.ucc.ie/en/inclusive/pilots/)

## Questions

Q1. How easily SensusAccess can deal with Mathematical notation?

A1. It is recommended to convert PDFs to Word and recreate any images of mathematical equations using the [Equation Editor in Word.](https://support.microsoft.com/en-us/office/write-an-equation-or-formula-1d01cabc-ceb1-458d-bc70-7f9737722702)

Q2. Can you batch process documents with SensusAccess?

A2. Unfortunately not. You could upload documents to Google Drive and convert to Google Docs, but some of the extra accessibility features of SensusAccess will be lost.

Q3. How accurate is SensusAccess?

A3. It is not always perfect. When converting a PDF to a Tagged PDF for example, if you select ‘Text-over-Image’ this may extract text from a logo and distort it. You may need to convert to Word and reformat some things.

**Note**: There is an option when scanning, to perform an OCR scan as well.