# Table of Contents

1 **Relativity Assisted Review**  
1.1 Latent Semantic Indexing  
1.2 Support Vector Machine learning  
1.3 Common project use cases  
1.3.1 Scenario 1: Review prioritization  
1.3.2 Scenario 2: Review all responsive items  
1.3.3 Scenario 3: Quick production  
1.3.4 Scenario 4: Identify the opposition productions' most relevant documents  
1.3.5 Scenario 5: QC a document set prior to production  

2 **Sample-Based Learning**  
2.1 Set project goals  
2.2 Perform Sample-Based Learning setup  
2.3 Prepare your reviewers  
2.4 Perform Sample-Based rounds  
2.5 Complete your Sample-Based Learning project (stabilization)  
2.6 Environment setup  
2.6.1 Sample-Based Learning system requirements  
2.6.2 Adding the Sample-Based Learning application  
2.6.3 Agent configuration  
2.6.4 Manually adding worker agents  
2.6.5 Fields  
2.7 Workspace setup  
2.7.1 Required workspace components  
2.7.2 Sample-Based Learning project checklist  
2.8 Sample-Based Learning document review  
2.8.1 Assigning round documents to reviewers  
2.8.2 Sample-Based Learning standards and protocol  

3 **Project setup**  
3.1 Creating a Sample-Based Learning project
3.1.1 Fields .................................................................................................................. 23
3.2 Console on the saved project layout ....................................................................... 27
  3.2.1 Project Home ........................................................................................................ 29
3.3 Viewing categorized and uncategorized document saved searches ......................... 30
3.4 Viewing round sample set document saved searches .............................................. 32
3.5 Viewing a round’s overturned documents (and seeds causing the overturn) .......... 33
3.6 Viewing a RAR project’s categorized documents in the Field tree ......................... 34
3.7 Modifying issue importance ..................................................................................... 34
3.8 Deleting a project ..................................................................................................... 35

4 Rounds ..................................................................................................................... 36
  4.1 Preparation before starting Sample-Based Learning rounds .................................. 36
  4.2 Types of rounds ....................................................................................................... 36
  4.3 Monitoring round progress ..................................................................................... 37
  4.4 Deleting a round ..................................................................................................... 39
  4.5 Control set round .................................................................................................... 40
    4.5.1 Executing a control set round ............................................................................ 41
    4.5.2 Reviewing documents for a control set round .................................................. 43
    4.5.3 Finishing a control set round ............................................................................ 43
    4.5.4 Reviewing Assisted Review reports after a control set round ......................... 44
  4.6 Pre-coded seed round ............................................................................................ 45
    4.6.1 Executing a pre-coded seed round ................................................................... 45
    4.6.2 Reviewing documents for a pre-coded seed round ........................................... 47
    4.6.3 Finishing a pre-coded seed round ................................................................... 48
    4.6.4 Reviewing Sample-Based Learning reports after a pre-coded seed round .......... 50
  4.7 Training round ....................................................................................................... 50
    4.7.1 Executing a training round .............................................................................. 50
    4.7.2 Training round document review ..................................................................... 53
    4.7.3 Finishing a training round ............................................................................... 53
    4.7.4 Reviewing Sample-Based Learning reports after a training round .................. 56
  4.8 Quality Control (QC) round .................................................................................. 56
    4.8.1 Executing a QC round .................................................................................... 56
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8.2 Reviewing documents for a QC round</td>
<td>58</td>
</tr>
<tr>
<td>4.8.3 Reviewing Sample-Based Learning reports during a QC round</td>
<td>58</td>
</tr>
<tr>
<td>4.8.4 Evaluating overturns and making corrections</td>
<td>59</td>
</tr>
<tr>
<td>4.8.5 Finishing a QC round</td>
<td>60</td>
</tr>
<tr>
<td><strong>5 Sample-Based Learning reports</strong></td>
<td>63</td>
</tr>
<tr>
<td>5.1 Report types</td>
<td>63</td>
</tr>
<tr>
<td>5.1.1 Designation reports</td>
<td>63</td>
</tr>
<tr>
<td>5.1.2 Control set reports</td>
<td>64</td>
</tr>
<tr>
<td>5.1.3 Issue reports</td>
<td>64</td>
</tr>
<tr>
<td>5.1.4 RAR project document reporting views</td>
<td>64</td>
</tr>
<tr>
<td>5.2 Viewing reports</td>
<td>64</td>
</tr>
<tr>
<td>5.3 Designation reports</td>
<td>65</td>
</tr>
<tr>
<td>5.3.1 Round Summary report</td>
<td>65</td>
</tr>
<tr>
<td>5.3.2 Rank Distribution report</td>
<td>70</td>
</tr>
<tr>
<td>5.3.3 Overturn Summary report</td>
<td>71</td>
</tr>
<tr>
<td>5.3.4 Project Summary report</td>
<td>76</td>
</tr>
<tr>
<td>5.4 Control set reports</td>
<td>78</td>
</tr>
<tr>
<td>5.4.1 Understanding richness, precision, recall, and F1</td>
<td>78</td>
</tr>
<tr>
<td>5.4.2 Control Set Statistics report</td>
<td>80</td>
</tr>
<tr>
<td>5.5 Issue reports</td>
<td>87</td>
</tr>
<tr>
<td>5.5.1 Issue Summary report</td>
<td>87</td>
</tr>
<tr>
<td>5.5.2 Designation-Issue Comparison report</td>
<td>89</td>
</tr>
<tr>
<td>5.6 RAR project document reporting views</td>
<td>94</td>
</tr>
<tr>
<td>5.6.1 Viewing control set documents</td>
<td>94</td>
</tr>
<tr>
<td>5.6.2 Viewing overturned documents</td>
<td>95</td>
</tr>
<tr>
<td>5.6.3 Viewing saved categorization results</td>
<td>98</td>
</tr>
<tr>
<td><strong>6 Errors and audit history</strong></td>
<td>100</td>
</tr>
<tr>
<td>6.1 Viewing errors</td>
<td>100</td>
</tr>
<tr>
<td>6.2 Retrying errors</td>
<td>102</td>
</tr>
<tr>
<td>6.2.1 Audit history</td>
<td>103</td>
</tr>
</tbody>
</table>
1 Relativity Assisted Review

Relativity Assisted Review (RAR) is a set of tools that helps you categorize your documents and automate the review process while minimizing the time your review team would otherwise spend coding irrelevant documents in your document set.

Read if you’re studying for the Assisted Review Specialist or Project Management Specialist exam

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

Note: Beginning in 9.5.370.136, Assisted Review projects are now called Sample-Based Learning projects.

View related recipes:

Imagine you're a member of a law firm representing a large energy company that becomes involved in litigation over an environmental matter and needs to quickly identify a list of relevant documents out of 1 TB of data. To complicate matters, there's only one attorney dedicated to the case. You're able to leverage Relativity's processing functionality to get the list down to 160,000 documents, but too many overly inclusive keywords like environment leave you stuck with too many non-responsive documents in the set for one person to review.

To deal with this situation, you create an Assisted Review project, the basic goal of which is to train the system to differentiate between responsive and non-responsive files related to this particular environmental matter and thus cull your data set down to a size that one person could feasibly review.

The attorney codes a random sample of 1,000 documents that were not previously reviewed, during which 6-7% of the documents are identified as responsive.

The same attorney then wants to test Relativity's accuracy by conducting a QC round with a statistical sample of documents from the default RAR project search containing categorized documents (based on a 95% confidence level and a 2.5% margin of error).

You're a litigation support specialist at a Relativity service provider, and the legal department of a large financial services company reaches out to you because the federal government is demanding that documents belonging to three key custodians be turned over quickly as part of an ongoing investigation.

This company is in a serious time crunch because the government agency’s attorneys then unexpectedly request documents from a fourth custodian, whom they believe is crucial to the
case. This doubles the size of the data they're required to review and produce, so they turn to you and you turn to Assisted Review.

You create a project that uses an Analytics index that includes the data of all four custodians. The project uses documents that were previously coded to expedite the training of the system. Relativity categorizes the document universe for prevalence, and Reviewers begin reviewing more documents to assist the system in deciding relevance.

In a Sample-based Learning project, you facilitate five training rounds on your new project and find that the turnover rate for non-responsive documents is low enough to make you confident that reviewers had identified all potentially responsive documents.

In an Active Learning project, reviewers are continuously provided documents of a certain rank. After reviewing a subset of documents, At the end of this project, you learn that less than 15% of the total documents in the universe needed review to produce accurate results in a limited time frame. The financial services company you're assisting can now easily comply with the federal government and give them what they need.

1.1 Latent Semantic Indexing

Sample-based Learning uses a type of text analytics called Latent Semantic Indexing (LSI). With LSI, you provide the Analytics index with sample text with which to identify concepts in your documents that are present but may not be obvious. A query can then map to these concepts.

The system then identifies similar documents based on context and concept rather than keyword. If apples and oranges occur together in many documents, the system identifies a concept that includes these two phrases. For example, if the word apples occurs in an email about oranges, LSI groups documents pertaining to oranges, rather than only documents about apples. LSI only uses text in its decision-making; numbers, symbols, and images are not considered.

After providing Relativity with example documents, the system conducts a categorization in which it matches a new document to this sample set for a given topic. Experts can identify documents relevant to a case so Relativity can then categorize new documents as relevant, not relevant, or pertaining to various issues based on the sample set.

1.2 Support Vector Machine learning

Active Learning uses a type of technology called Support Vector Machine Learning (SVM). With SVM, you don't need to provide the Analytics index with any training text. The system learns from your reviewers and constantly updates the model.

SVM technology creates a multi-dimensional space from coding decisions and divides the relevant from non-relevant examples through what is known as a “hyperplane”. The system predicts uncoded documents against the coded examples so that each document's rank reflects a document's probability of being relevant. This differs from Latent Semantic Indexing (LSI) which is also composed of a multi-dimensional space, but uses a nearest-neighbor approach to predict documents. SVM predicts uncoded documents' relevance based on their closeness to coded examples. Each document is given a rank from 0 to 100. The closer to 100, the more confident the model feels that the document is likely to be relevant.
1.3 Common project use cases

How you proceed with your project depends on your case and the risks involved in the production of privileged or non-responsive material. Using the Assisted Review layout, reviewers can validate the system-categorized values.

Generally, cases fall into one of the common scenarios included in this section. Note that these scenarios represent suggested workflows and only provide an overview of the process. If you need assistance with a specific matter, please contact solutions@relativity.com.

1.3.1 Scenario 1: Review prioritization

In this scenario, attorneys may want to review the entire document population. The goal is to get the most important documents to the review team as soon as possible. The remaining documents will still be reviewed, but perhaps later by a review team at a lower billing rate. This process allows them to determine resources after a couple of rounds. Prioritization projects typically don’t require as many rounds as other types of projects, because all documents are eventually reviewed.
1.3.2 Scenario 2: Review all responsive items
In this scenario, the review team manually reviews all responsive documents but trusts the system based on acceptable error rates for the non-responsive population. The non-responsive documents are set aside and aren't reviewed. Privilege is not a major concern for this group. Using search terms across responsive items for privilege is an acceptable method of privilege review.

1.3.3 Scenario 3: Quick production
In this scenario, documents need to be produced in a very short time frame. It isn't a strong concern whether the production is over-inclusive, meaning it can include a few non-responsive items. In addition, privilege screening isn't typically a major concern for this scenario.

The basic goal of this approach is to achieve a low uncategorized percentage along with a low estimated defect percentage before finalizing the project and proceeding to an accelerated production.

1.3.4 Scenario 4: Identify the opposition productions’ most relevant documents
When the other side of a litigation produces documents to you, there is an inclination to presumptively treat the entire production as responsive. As such, Assisted Review projects of this nature are designed to locate the documents that are most beneficial to your case.

1.3.5 Scenario 5: QC a document set prior to production
In this scenario, the project manager leverages the technology to assist with QC of an existing manual review project. It’s a conservative and very useful method to learn if any documents have been missed or coded inconsistently.
2 Sample-Based Learning

Sample-Based Learning uses a repetitive process to group and code documents. The process takes a small group of manually-coded documents and treats them as a representation of the entire document set. Based on the text in that group of documents, Sample-Based Learning categorizes all the documents in your workspace.

Read if you're studying for the Assisted Review Specialist or Project Management Specialist exam

The content on this site is based on the most recent monthly version of Relativity, which contains functionality that has been added since the release of the version on which Relativity's exams are based. As a result, some of the content on this site may differ significantly from questions you encounter in a practice quiz and on the exam itself. If you encounter any content on this site that contradicts your study materials, please refer to the What's New and/or the Release Notes on the Documentation site for details on all new functionality.

Note: Beginning in 9.5.370.136, Assisted Review projects are now called Sample-Based Learning projects.

The following diagram outlines the basic Sample-Based Learning workflow.

This topic contains the following sections that outline everything you need to get started with Sample-Based Learning:

1. Set project goals below
2. Perform Sample-Based Learning setup on the next page
3. Prepare your reviewers on the next page
4. Perform Sample-Based rounds on page 11
5. Complete your Sample-Based Learning project (stabilization) on page 12

2.1 Set project goals

Before you begin a Sample-Based Learning project, we suggest you think about whether the project is a good fit for Sample-Based Learning. Sample-Based Learning is centered on the concept of training the
system so that it learns how to interpret uncategorized documents. The system learns best from
documents that are good examples. To be good examples, documents should have rich text with lots of
concepts, not just numbers.

Consider also what constitutes a responsive document. If, for instance, responsiveness hinges on a name
or a date, that is likely not enough for Sample-Based Learning because there are no concepts to learn,
only absolutes. Successfully completing a Sample-Based Learning project requires you to spend a little
time at the beginning to determine whether Sample-Based Learning is the best way to proceed.

Every Sample-Based Learning project has specific needs, goals and deliverables. This checklist is meant
to be customizable to fit the needs of each project, but useful as a guide to the most commonly required
items.

1. Ensure the document set you plan to use is a good population for Sample-Based Learning:
   - Minimum 50k records with text
   - Concept rich files (not primarily numbers)
   - Issue or privilege coding is in a separate field or not part of Sample-Based Learning workflow

2. Make sure your timeline and goals are set. The stakeholders should discuss goals and timelines
   prior to beginning a Sample-Based Learning project so that clear deliverables are established.
   - Level of Precision, Recall, and F1 determined
   - Manual review plan decided (i.e., all docs categorized as Responsive; privilege screen only)
   - Production plan in place

### 2.2 Perform Sample-Based Learning setup

1. First set up your environment for the Sample-Based Learning project. See [Environment setup on
   page 12](#).

2. Set the Tab Visibility workspace security permission for Sample-Based Learning. See the Admin
   Guide for more information.

3. Next, set up your Sample-Based Learning workspace. See [Workspace setup on page 17](#).

### 2.3 Prepare your reviewers

Make sure your reviewers are prepared. Reviewer preparation is key to success. A Sample-Based
Learning project is not like other document coding attorneys may have done, so use all the tools available
to be sure everyone is trained in Sample-Based Learning protocols.

1. [Sample-Based Learning for End Users webinar](#) has been viewed.

2. Sample-Based Learning Reviewer Protocol has been distributed and discussed. For more
   information, see Recipes.
2.4 Perform Sample-Based rounds

1. Create the Sample-Based Learning project based on the goals you've set.

2. (Optional) Control set round - Identify a representative sample group of documents as your control set and have reviewers code these documents.

   A control set is used to automatically determine precision and recall and F1 for your project using Sample-Based Learning reporting.

3. Training round - Identify a sample group of documents in your data set to train the system with, and assign this to reviewers to code this training sample group and set the example documents.

   **Note:** Alternatively, if reviewers have already coded representative documents per Sample-Based Learning protocol, you can use the group of documents as a pre-coded seed round to train the system with.

4. Submit the round sample documents to the system by finishing the round in order to categorize your entire data set.

   Each document in your searchable set is categorized based on the closest example document.

   **Note:** You may repeat the prior two steps until the system has categorized a certain desired percentage of documents.

5. QC round - Sample a group of documents categorized by the system by creating a QC round, and then have reviewers review and code this sample set of documents to quality control (QC) the system.

6. Before finishing the QC round, perform overturn analysis using Sample-Based Learning reporting to find seed documents that created the most overturns. Work with reviewers to ensure that the seed documents are correctly defined. After making fixes, finish the round.

   **Note:** Throughout the process, analyze your progress using the Sample-Based Learning reporting and then verify whether you’re finished with the process or need to complete another iteration.

7. Continue this process until the project reaches a stable point as determined from your goals and reporting.

**Note:** Sample-Based Learning is centered on the concept of training the system. Sample-Based Learning trains the system by learning how to interpret uncategorized documents. This equips the system to successfully categorize others with a high level of confidence and accuracy. Reviewers continue to train the system by manually reviewing documents and assigning them to categories.
2.5 Complete your Sample-Based Learning project (stabilization)

Planning in advance will ensure a successful wrap up. Ensuring that all tasks are complete is important for the client’s satisfaction as well as defensibility. The following should be satisfied before you can consider a project complete.

- Project goals met
- Precision/recall
- Stabilization achieved
- Manual review under way
- Production complete

Once you reach your goal, you can continue to the next phase of review. After your project reaches stabilization and the overturn rate percentage of change in responsiveness stabilizes, you can take the values determined by Sample-Based Learning to proceed towards production or organization of documents for case work. This is the time to start creating these document groupings. The path you take is dependent on your project goals.

Consider the following post-project completion tasks:

- Executing searches to find responsive documents and include family items
- Manually reviewing documents that didn’t get a categorization value and aren’t part of the responsive family group
- Reviewing responsive items for privilege
- Spot-checking non-responsive items
- Organizing case files around relevance
- Creating witness binders around issues

2.6 Environment setup

You must first verify that your system and workspace meet the necessary standards, and then perform the required installation and configuration steps to successfully run a Sample-Based Learning project.

2.6.1 Sample-Based Learning system requirements

Before you can begin to use Sample-Based Learning you must have the following:

- Relativity 7.5 or higher installed, and Sample-Based Learning added to the workspace. See Adding the Sample-Based Learning application on the next page for more information.
- The Tab Visibility permission for RAR. See the Admin Guide for more information.
- To use email notifications, the SMTP server must be configured; you can do this through the Relativity SMTP Configuration script, which you can run to set up the available SMTP fields in the kCura.Notification section of the Instance setting table. See the Relativity SMTP configuration script for more information.
2.6.2 Adding the Sample-Based Learning application
To install Assisted Review in your workspace, import the Assisted Review application from the application library. To do this, you must have the appropriate admin rights.

To install the Assisted Review application in your workspace:

1. Navigate to the Relativity Applications tab.
2. Click New Relativity Application.
3. Select Select from Application Library.
4. Click on Choose from Application Library.
5. Select Assisted Review, and click OK.
6. Click Import.

Once the import completes, Relativity creates an Assisted Review tab, from which you can use Assisted Review in your workspace. You may need to refresh your browser to see this new tab.

2.6.3 Agent configuration
Sample-Based Learning functionality depends on two types of agents:

- Assisted Review Worker - does the heavy lifting. The more worker agents there are, the faster some Assisted Review processes are. However, there is a limit to how many are useful, and using too many may cause unnecessary load on some Relativity setups.
- **Assisted Review Manager** - does some organization and initialization work, specifically overseeing distribution of processes to the worker agents. This agent works on all projects regardless of Resource Group. It does no intensive work. You can have only one manager agent per environment hosting Assisted Review, as opposed to one manager agent per agent server. If you install more than one manager agent in the environment, Assisted Review fails to operate in Relativity.

Note the following about RAR agents:

- One worker and one manager agent are installed automatically during the initial installation of Assisted Review.
- Never put RAR agents on a web server.
- Always disable your RAR agents when upgrading Relativity, or else you will encounter issues with your RAR project(s) in your new Relativity environment.

### 2.6.3.1 Agent recommendations

The following guidelines are a good place to begin when setting up your Relativity environment for Assisted Review:

<table>
<thead>
<tr>
<th>Environment Size</th>
<th># of Worker Agents</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier 1 (Single Server Deployments)</td>
<td>2</td>
<td>More than two Assisted Review agents may introduce system wide performance and stability related issues.</td>
</tr>
<tr>
<td>Tier 1 - Tier 2</td>
<td>&lt; 10</td>
<td>Dependent on available resources.</td>
</tr>
<tr>
<td>Tier 3</td>
<td>10 - 30</td>
<td>When running many Assisted Review projects on many workspaces, going past ten agents is beneficial. Work with Client Services to determine your needs. If you have more than ten Worker agents, you can raise the run interval of the additional agents from .5 seconds to 5 seconds.</td>
</tr>
</tbody>
</table>

See the System Requirements guide for more information.

Note the following about categorization in Sample-Based Learning:

- The most resource-intensive Sample-Based Learning processes are categorization and saving categorization results.
- Categorization does not utilize Assisted Review agents; it uses the Analytics Categorization Manager.
Assisted Review requires the use of an Analytics Categorization Manager. There should be no more than two Analytics Categorization Manager Agents per Relativity resource pool. See the Agents Guide for more information regarding the Analytics Categorization Manager.

In consultation with Relativity Client Services, you may want to consider adding worker agents if:

- You have many more Assisted Review projects than you have Assisted Review Worker agents.
- You commonly save the categorization results of multiple projects at the same time.
- The saving of categorization results takes a long time.

2.6.3.2 Relevant instance setting table values
RAR uses the following instance setting table values to retrieve data:

- DefaultQueryTimeout
- LongRunningQueryTimeout

The following EDDS instance setting table values control Assisted Review error retries:

- Retries
- RetryWaitTime

For more information, see the Instance setting guide.

2.6.4 Manually adding worker agents
Depending on your environment setup, adding more worker agents may increase performance of Assisted Review. To add worker agents:

1. Click your name in the upper right corner of Relativity and click Home.
2. Navigate to the Agents tab.
3. Click New Agent.
4. Complete the fields on the agent form. See Fields below for more information.
5. Click Save. You receive a confirmation that the agent was successfully added.

2.6.5 Fields
The agent form contains the following fields:
- **Agent Type** - the type of agent you're adding. Click to display a list of available agents. Select **Assisted Review Worker** and click OK.

- **Number of Agents** - the number of agents you want to add. Environments hosting Assisted Review should have multiple worker agents installed to avoid performance issues when executing a project. Even if only a single project is running in the environment, you should have at least two worker agents installed.
- **Agent Server** - the server where you want to add the agent(s). Click to display a list of available servers. Select the server and click OK.

- **Run Interval** - the number of seconds between each time the agent checks for available jobs. When you select Assisted Review Worker as the agent type, this value automatically defaults to 0.5. Changing this value impacts performance.

- **Logging level of event details** - determines what events the agent logs. For Assisted Review, it is recommended that you keep this at the default value of Log critical errors only.

- **Enabled** - determines whether the agent is enabled once you add it. For Assisted Review, it is recommended that you keep this at the default value of Yes.

See the Agents Guide for more information.

### 2.7 Workspace setup

A Sample-Based Learning project is an object in which you define the document set to use as the basis for categorizing a larger set of documents, as well as the responsiveness fields used by reviewers to code the set of documents that you want Relativity to recognize so that it can correctly categorize responsive and non-responsive documents going forward using that set as the basis. Once you create a project, you run rounds on it to further specify how documents should be categorized.

#### 2.7.1 Required workspace components

A Sample-Based Learning project uses the following components, so you'll need to create them before you can create a project. Even if these items already exist in the workspace, you may want to create a new instance of each specifically for your new project.

Before reviewing for a round, all the document coding fields set below during project creation should be available on a document layout (e.g., Designation, Example Responsiveness). See Layouts in the Admin Guide for more information.

**Note:** You cannot use Family propagation in a Sample-Based Learning project. Furthermore, if you use document skip in a Sample-Based Learning project and you intend to code the entire sample set, you must check to make sure that all sample set documents get coded.

The following graphic depicts the components required to save a new project. If any of the items below are missing from your workspace, you won't be able to create and save the project.
2.7.1.1 Analytics index
You must have an Analytics index that includes the documents to be used in the project.

See the Admin Guide for more information.

**Note:** The Analytics index you use for your project must be active and have queries enabled for your project to function properly.

2.7.1.2 Fields
Necessary fields for setting up and running a project include:

- **Designation field** – this single choice field is limited to two choices, for example, Responsive and Not Responsive. You will get an error if you attempt to create a project that has too many choices on this field, or if no choices have been created.

- **Use as Example field** – this Yes/No field is used by reviewers to designate a document as an example to use for categorizing other documents.

- **Text Excerpt field** - this long text field holds the text excerpts you apply to coding layouts during manual review (if you are running a project that will use text excerpting). This enables reviewers to select the relevant text during document coding to enhance the training of the system by including the relevant portion of the document as an example.
Note: If you excerpt text and save it in the field for a document, that text is always an example regardless of the Use as Example selection. To include the entire document and the text excerpt as two separate examples, select Use as Example. Clear the Use as Example checkbox to only include the text excerpt as an example.

- **Key Issues field** (Optional) - Single choice field used for issues (if you are running a project that will categorize documents for key issues). See Modifying issue importance on page 34 for more information on how to weight issue importance.

- **Key Issues Text Excerpt field** (Optional) - Long text field used for text excerpting (if you are running a project that will use text excerpting). This enables reviewers to select the relevant text during document coding to enhance the training of the system by including the relevant portion of the document as an example. See Modifying issue importance on page 34 for more information on how to weight issue importance.

### 2.7.1.3 Saved search

**Project saved search** - includes documents to be categorized that do not include documents that cannot be categorized (e.g., documents must have text). See the Admin Guide for more information.

### 2.7.2 Sample-Based Learning project checklist

You may want to use the following checklist to make sure that you have all the items needed to run a Sample-Based Learning project:

<table>
<thead>
<tr>
<th>Sample-Based Learning component</th>
<th>Completed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Install the Assisted Review application</td>
<td></td>
</tr>
<tr>
<td>Complete Analytics index build</td>
<td></td>
</tr>
<tr>
<td>Set up the following required project fields:</td>
<td></td>
</tr>
<tr>
<td>- Designation field - single choice</td>
<td></td>
</tr>
<tr>
<td>- Choices: Responsive, Not Responsive</td>
<td></td>
</tr>
<tr>
<td>- Use as Example field - Yes/No</td>
<td></td>
</tr>
<tr>
<td>- Text Excerpt field - Long text</td>
<td></td>
</tr>
<tr>
<td>(Optional) Set up the following fields (when categorizing for issues):</td>
<td></td>
</tr>
<tr>
<td>- Key Issues field - Single choice</td>
<td></td>
</tr>
<tr>
<td>- Key Issues Text Excerpt field - Long text</td>
<td></td>
</tr>
<tr>
<td>Create any necessary reviewer views</td>
<td></td>
</tr>
<tr>
<td>Make all the document coding fields set during project creation available on a document layout.</td>
<td></td>
</tr>
<tr>
<td>- Designation, Use as example, Text excerpt field, etc.</td>
<td></td>
</tr>
</tbody>
</table>
### Sample-Based Learning component

<table>
<thead>
<tr>
<th>Sample-Based Learning component</th>
<th>Completed?</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(Optional)</em> Perform the following step especially when categorizing for issues:</td>
<td></td>
</tr>
<tr>
<td>- Cluster all documents in a Sample-Based Learning universe</td>
<td></td>
</tr>
<tr>
<td>Perform document culling to remove documents that won’t work well with Analytics indexes:</td>
<td></td>
</tr>
<tr>
<td>- Eliminate documents that do not contain text</td>
<td></td>
</tr>
<tr>
<td>- Run the extracted text size script to review which documents are extremely low on text</td>
<td></td>
</tr>
<tr>
<td>- If documents are clustered, pull UNCLUSTERED documents from the project.</td>
<td></td>
</tr>
<tr>
<td><em>(Optional)</em> You can also use the following additional operations to cull documents that may not be important to the case:</td>
<td></td>
</tr>
<tr>
<td>- Email threading</td>
<td></td>
</tr>
<tr>
<td>- Keyword culling</td>
<td></td>
</tr>
<tr>
<td>- Search terms reports</td>
<td></td>
</tr>
<tr>
<td>Create any necessary saved searches:</td>
<td></td>
</tr>
<tr>
<td>- Project saved search - documents to be categorized that do not include documents that cannot be categorized (e.g., documents must have text).</td>
<td></td>
</tr>
</tbody>
</table>

## 2.8 Sample-Based Learning document review

### 2.8.1 Assigning round documents to reviewers

Reviewers can begin coding documents at any point after the project has a status of RAR **Sample Set created**. To give reviewers access to documents designated for a specific project round, project managers can create views or batches of documents based on the RAR sample set field. As document review progresses, the [progress bar](progress-bar) on the project layout is updated accordingly.
When the **Sample Set Size** and **Reviewed Count** fields contain the same value, it means that all documents in the sample set were coded and the document review is complete.

Additionally, when you are creating a new round, you may choose to automatically create batches to assign round sample documents to reviewers when a round is initiated by selecting Yes for the Automatically create batches field. You can also set the **Maximum Batch Size** to specify the maximum number of documents in a batch. You would then assign those batches to your designated RAR reviewers as you would normally for document review.

### 2.8.2 Sample-Based Learning standards and protocol

It’s important for reviewers to be diligent when deciding which documents are good examples for machine learning. See [Relativity Assisted Review Reviewer Protocol](#) for more detailed information on how reviewers should code documents during their review.

**Note:** Sample-Based Learning does not support mass edit. If reviewers attempt to code documents in the sample set using mass edit, Sample-Based Learning does not update the Reviewed Count field or the overturn information. Sample-Based Learning also does not capture the seed reviewer of the documents for use in subsequent rounds. Categorization, however, includes mass edited documents as examples.

### 2.8.2.1 Text excerpting considerations

Reviewers can apply text excerpts during coding:
1. Highlight a section of text.
2. Right-click to launch the context menu.
3. Select Assisted Review and then the appropriate project.
4. Select the corresponding field in the layout.
5. Verify that the appropriate text from the document is copied to the layout and click **Save**.

Note the following about the context menu in the viewer:

- The Assisted Review option is only available in the context menu if there is a project created in the workspace.
- A project is only available to select from the context menu if either the Designation excerpt field or Key issue excerpt field is set for the project.
- The excerpted text option for designation is only available in the context menu if the Designation excerpt field is set for the project.
- The excerpted text option for issues is only available in the context menu if the Key issue excerpt field is set for the project.
- Only users that have the Local Access Permission on Document for the workspace see the Assisted Review context menu.

**Note:** If you excerpt text and save it in the field for a document, that text is always an example regardless of the Use as Example selection. To include the entire document and the text excerpt as two separate examples, select Use as Example. Clear the Use as Example checkbox to only include the text excerpt as an example.

**Note:** Relativity displays a warning message if too little text is selected as an excerpt. By default this message displays if fewer than 120 characters are selected. You can update this value using the MinimumExcerptTextLength entry in the Instance setting table.
3 Project setup

A Sample-Based Learning project is an object in which you define the document set you plan on using as the basis for your predictive coding job, as well as the fields you want Relativity to recognize as responsive so that it can differentiate between responsive and non-responsive documents from that set. Once you create a project, you run rounds on it to further determine how documents should be coded.

3.1 Creating a Sample-Based Learning project

To create an Assisted Review project:

1. Click the Sample-Based Learning tab and then click the Projects tab.
2. Click New Assisted Review Project.
3. Complete all required fields on the project settings page. See Fields below.
4. Click Save. This stores the new project.

When you create a new project, Sample-Based Learning creates a saved search folder for the project containing the saved search you specified in the project console for the Documents to be Categorized field.

Notes:
- A saved search folder designated for the sample sets under the project saved search folder will be created when you create your first round and generate a sample set. This folder will contain a saved search for the sample set of documents that are generated for each round you create. Sample-Based Learning uses the sample set saved search as the data source when you are automatically creating batch sets for reviewers. See Rounds on page 36 for more information.
- The default category searches (e.g., Categorized, Not Responsive, Responsive, and Uncategorized) are automatically created in the project-specific saved search folder when you finalize your first training round and categorize documents for the first time. After documents are categorized, the documents in each search set will by default have a minimum coherence score of .7 in relation to the example documents that categorized them. See Viewing categorized and uncategorized document saved searches on page 30.

3.1.1 Fields

The project settings layout contains the following fields:
- **Project name** - the name of the project. Make this unique to differentiate it from other projects in the workspace.

- **Project prefix** - the group of letters used to identify the project in various reports and layouts. For example, P01. This is automatically populated for you when you go to create the project, but it is editable. You can't use some special characters, such as / and \, in the project name or prefix or you will receive an error.

- **Project status** - an automatically populated read-only display of the project's current state. Status values and their meanings are covered later on this page.

- **Project description** - a text box for entering a description for the project.

- **Analytics index** - the index used to categorize documents during the project. Click to display a list of all Analytics indexes. If there are no active indexes in the workspace, or if you want to create an index specifically for the purposes of this project, you must navigate to the Search Indexes tab and create an index before beginning the project. This index does not need to be active when you add it to the project; however, the index must be active when you are finishing a round. If the index becomes inactive at the time you are finishing a round, you get an error.

- **Documents to be categorized** - the saved search containing all the documents included when categorizing the project.

  **Note:** The saved search you use should exclude documents that cannot be categorized.

- **Designation field** - the field used to code each document. Click to display a pop-up picker of available single choice fields. If the desired field is not available, you must create it in the Fields tab. See the Admin guide for more information. This value is typically a standard designation or responsiveness field. If you have multiple projects in the same workspace, you must use a different Designation field for each project. This single choice field must contain **two** choices; more than two choices results in an error when you attempt to save the project. Once you save the project, you can't add, delete, or rename a choice for the Designation field. This is to ensure that all statistics, graphs, and reports are calculated and render correctly. If you delete the project, the object rule that disallows you from adding or removing a choice from the designation field is also deleted.

- **Positive choice for designation** - select the positive choice radio button for the Designation single choice field that you selected above. The positive choice for this field refers to the set of documents you are trying to find (e.g., the responsive or relevant documents). Sample-Based Learning uses this selection when calculating precision and recall for the control set. This field is required and populated with two choices after you select a Designation field above. You can't edit this field after you save the project.

- **Designation excerpt field** - used for applying text excerpts from documents to the coding layout during manual review for designation. Using excerpts while manually reviewing enhances the training of the system because the relevant part of the document is included as an example. The field you select here is available in the viewer’s context menu for reviewers’ coding. Click to display a list of available long text fields.
Use as an example field - used to indicate which documents are good examples for training the system. This field is set to yes by default on all documents in the sample set, except for control set rounds, to indicate that those documents are examples. The reviewer should de-select this field on a document if the document is not a good example. Doing this prevents poor examples from being included in the Analytics examples. You can’t use an example field across projects.

Note: For more information on adding example documents, see the Analytics Guide.

Key issue field - the field reviewers use to code documents with a single issue. Click to display a list of available single choice fields. You can’t edit this field once you save the project. A second categorization set is created and run for this issue field once you save the project. If you don’t add an issue field when initially creating the project but you want to later, you can only add an issue coding field between rounds. You can use the same Key issue field across multiple projects. See Modifying issue importance on page 34 for more information on how to weight issue importance.

Key issue excerpt field - used for applying text excerpts from documents to the coding layout during manual review for issues. Using excerpts while manually reviewing enhances the training of the system because the relevant part of the document is included as an example. The field you select here is available in the viewer’s context menu for reviewers’ coding. Click to display a list of available long text fields. See Modifying issue importance on page 34 for more information on how to weight issue importance.

Confidence level - the probability that the rate in the sample is a good measure of the rate in the project universe. This value is the default for the confidence level field when you start a round. The choices are: 99%, 95%, and 90%. Selecting a higher confidence level requires a larger sample size.

Sampling type - the method used to create the sample set. The sample set is the randomly-selected group of documents produced by Sample-Based Learning to be used for manual review as a means of training the system. Select one of the following:

- Statistical sampling - creates a sample set based on statistical sample calculations, which determines how many documents your reviewers need to code in order to get results that reflect the project universe as precisely as needed. Selecting this option makes the Margin of error field required.

- Percentage - creates a sample set based on a specific percentage of documents from the project universe. Selecting this option makes the Sampling percentage field required.

- Fixed sample size - creates a sample set based on a specific number of documents from the project universe. Selecting this option makes the second Fixed sample size field required.

Note: To execute sampling, Sample-Based Learning uses a randomization algorithm called the Fisher-Yates shuffle, which guarantees an efficient and unbiased result.

Margin of error - the predicted difference between the observed rate in the sample and the true rate in the project universe. This is the amount of random sampling error you can allow for the project when you select a sampling type of statistical sampling. The options are +/- 0.5%, +/-1.0%, +/-
1.5%, +/-2.0%, +/-2.5%, and +/-5.0%. Selecting a lower margin of error requires a larger sample size.

- **Sampling percentage** - the percentage of the universe you want to act as your sample. Enter any percentage between 1-100% when you select percentage as the sampling type.

- **Fixed sample size** - the number of documents out of the universe that you want to act as your sample when you select fixed samples size as the sampling type.

- **Automatically create batches** - determines whether or not batch sets and batches are automatically created for the project's sample set(s) to expedite review kickoff. Selecting Yes here makes the Maximum batch size field below required. The batch set and batches created from this project are editable after you create the project. By default, this field is empty. The value you select here appears as the default value when you are starting the first round.

- **Enter email addresses** - an optional text box where you list the email addresses of all recipients you want to receive notifications when various parts of the project have completed. Separate email addresses with a semi-colon. Email notifications are sent if the project encounters an error and after the following parts of the project have completed:
  - Sample set creation
  - Manual document review of the sample set
  - Categorization
  - Saving categorization results

3.1.1.1 Error scenarios when saving or editing a project
To preserve the unique settings of each project in the workspace, Relativity prevents you from performing the following actions. If you attempt to do any of these, you get an error:

- Save a new project under the name of an existing project.
- Save a new project with an existing project prefix or a prefix already used on a Categorization Set. For example, if you deleted a Sample-Based Learning project but did not delete the associated Categorization Sets and try to reuse the same prefix, you get an error.
- Save a project with a Document coding field that does not contain two choices.

3.2 Console on the saved project layout
The Sample-Based Learning console sits on the right side of the saved project layout. Upon first saving the project, limited options are available to select until the project creates the necessary Analytics categorization sets.

Click **Refresh Page** to update the console.
The console provides the following options:

- **View Project Home** - takes you to the project home layout, where you can view the project status, progress, and a list of all rounds in the project. For details, see [Project Home on the next page](#).

- **View Project Settings** - takes you to the project settings layout, where you can view, edit, or delete the settings you specified when you created the project.

- **Start Round** - allows you to start a new round for the project by selecting the appropriate round and sampling settings.

- **Finish Round** - marks the completion of the round at its current state and allows you to categorize and save results, depending on how many rounds you’ve completed and where you are in the project. This changes to Finish Round only after the round has been started.
View Round Summary - launches a graphical summary of the project categorization results by round.

View Rank Distribution - launches a graphical summary of the percentage of documents categorized for each designation choice in rank ranges.

View Overturn Summary - launches a graphical summary of overturned documents and percentages per round. This is only available after the second round has started.

View Project Summary - launches a report that provides a consolidated set of quality metrics from the entire project so that you can see the state of the project based on the last round completed. This includes manually coded documents, project categorization results, project overturn results, and control set statistics.

View Overturned Documents - launches details on documents that were manually coded by a reviewer differently from the value the system applied. This report is only available after the second round is started.

View Saved Categorization Results - launches details on all saved categorization results for issues and designation. This report is only available if results were saved while finishing a previous round.

View Issue Summary - launches a graphical summary of what percentage of documents in the project were coded with what issues.

View Designation-Issue Comparison - launches a graphical representation of documents’ designation vs. issue categorization results.

View Control Set Statistics - launches a graph and set of tables breaking down categorization results for all documents selected for a control set round, as well as data on precision, recall, and F1. This is only available if you have completed a control set round.

View Control Set Documents - launches a view that reflects how all documents in the control set were categorized each round. This is only available if you have completed a control set round.

Retry Errors - kicks off an attempt to retry errors encountered during the project. If the error occurs on a categorization set, this option is disabled and you have to go to the Categorization Set to resolve the error.

View Errors - takes you to a layout containing details on all errors encountered.

View Audit History - takes you to a layout containing all audited actions executed during the project.

Refresh Page - updates the page to its current status. After first saving the new project, you must click this at least once in order to enable the Start Round button. Refreshing the page after you start a round also allows you to see the current state of the round’s review progress in the round list.

For more information on reports, see Sample-Based Learning reports on page 63.

3.2.1 Project Home

When you navigate to the Project Home layout from the Sample-Based Learning console, you can view the Sample-Based Learning Round object, which provides a number of fields to help you understand the progress of your project.
- **Round name** - the name given to the round.
- **Round type** - the type of round: Training, Quality Control, Pre-coded Seeds, or Control Set.
- **Round description** - the text, if any, entered to describe the round.
- **Saved search name** - the name of the saved search used in the round.
- **Sampling type** - the sampling type used for the round.
- **Sample set size** - the number of documents in the round's sample set.
- **Review count for designation** - number of documents that have been coded with the designation field.
- **Seed count** - the number of documents in the round's sample set that are being used to train the system. This includes documents that have been excerpted.
- **Eligible sample docs** - the number of documents in the round eligible to be included in the sample set.
- **Docs in round** - the total number of documents in the round.
- **Docs in project** - the total number of documents in the project list.

### 3.3 Viewing categorized and uncategorized document saved searches

When you create your first project, Sample-Based Learning creates an unsecured RAR Saved Searches folder. For every project you create, Relativity creates a project-specific folder beneath the RAR Saved Searches folder containing the saved search you specified in the project console for Documents to be Categorized. When you create your first round, a RAR sample sets folder is created that will contain saved searches for each round’s sample set of documents that are batched out to reviewers.

After the first categorization occurs during the round finish process, the category saved searches are automatically created in the project-specific folder (e.g., saved searches that return Categorized, Not Responsive, Responsive, and Uncategorized documents in the project universe).

Note the following:

- You can add a search to the RAR Saved Searches folder, and you can select it when you start a round.
- Saved searches inherit security from the folder they are in. The RAR Saved Searches folder is unsecured by default.
- A saved search will still be created if there are no documents for a designation choice.
- If the system cancels the round, no saved search is created.

**Note:** All automatically-created RAR saved searches contain the phrase *RAR created* in the Keywords field to distinguish them from other searches.

All of the automatically-created searches listed below include the saved search criteria for your project's original saved search that you specified in the Documents to be Categorized field on the project console. (i.e., they’re all rooted in your project's original document universe).

Each search contains the following as the first field condition.
• Field: (Saved Search)
• Operator: Document is in
• Value: <Your project saved search>

You can select these searches for future rounds of the project in the **Saved search for sampling** field when you’re starting the round. The automatically-created saved searches are:

• **<Project Saved Search>** - this is the saved search you specified in the Documents to be Categorized field on the project console. The conditions/criteria are:
  • Field: (Saved Search)
  • Operator: Document is in
  • Value: Your project saved search

• **<Project Saved Search>** - Categorized - returns all categorized documents; this focuses the round’s sampling to categorization results for the purposes of QC. This search includes the <Project Saved Search> criteria and the following additional criteria:
  • Field: Categories - <Project Prefix> RAR Designation Cat. Set
  • Operator: these conditions
  • Value: Categories <Project Prefix> RAR Designation Cat. Set; is set

• **<Project Saved Search>** - Not Responsive - returns all documents that were categorized as not responsive during designation coding. This search includes the <Project Saved Search> criteria and the following additional criteria:
  • Field: Categories - <Project Prefix> RAR Designation Cat. Set
  • Operator: these conditions
  • Value: [Categories - <Project Prefix> RAR Designation Cat. Set] is any of these: Not Responsive

• **<Project Saved Search>** - Responsive - returns all documents that were categorized as responsive during designation coding. This search includes the <Project Saved Search> criteria and the following additional criteria:
  • Field: Categories - <Project Prefix> RAR Designation Cat. Set
  • Operator: these conditions
  • Value: [Categories - <Project Prefix> RAR Designation Cat. Set] is any of these: Responsive

• **<Project Saved Search>** - Uncategorized - returns all uncategorized documents and focuses the round’s sampling to those documents that haven’t been categorized yet. This search includes the <Project Saved Search> criteria and the following additional criteria:
  • Field: Categories - <Project Prefix> RAR Designation Cat. Set
  • Operator: not these conditions
  • Value: Categories <Project Prefix> RAR Designation Cat. Set; is set
3.4 Viewing round sample set document saved searches

Sample-Based Learning automatically creates a saved search for each round’s sample set in the project. Sample-Based Learning then uses those searches as data sources when automatically creating batch sets for reviewers. Sample-Based Learning puts sample set saved searches into the RAR Sample Sets subfolder of the project’s saved searches folder.

![Sample set saved searches](image)

*Note:* Sample set saved searches do not show up as choices in the Saved Search for Sampling field on the Start Round layout.

A sample set saved search is identified as `<Round Name> Sample Set` and contains the following criteria:

- **Field:** RAR Sample Set
- **Operator:** these conditions
- **Value:** [RAR Sample Set] is any of these: `<Round Name>`

*Note:* The RAR Sample Set folder and the searches it contains are deleted when the project is deleted. An individual sample set saved search is deleted when the round that created it is deleted.
3.5 Viewing a round's overturned documents (and seeds causing the overturn)

When you save a project, Relativity creates a multi-choice Document field called RAR overturn status - <Project prefix> that identifies overturned documents and the seed documents and excerpts that caused overturns per round. Each round that you start in the project automatically creates choices for overturned document, seed document, and seed excerpt, except for the first round, which can't contain overturns. You can select these in the field tree and/or create a view or a saved search with these sets as criteria.

These fields are valuable because:

- They make it easy to aggregate the documents found in the Overturned Documents report on the project, which reviewers might not have access to view.

Note the following about the seed excerpt choice in the overturn status field:

- For every overturn caused by a seed excerpt, the seed document containing that excerpt is tagged with the seed excerpt choice for the round in which the overturn occurred.
If you don’t select a Designation excerpt field for the project, the seed excerpt choices are not created.

The seed excerpt choice is still created if there were no overturns caused by seed excerpts.

### 3.6 Viewing a RAR project’s categorized documents in the Field tree

You can also view documents that have been categorized by RAR for both issues and designation in the field tree.

RAR field tree categories are identified as **Categories - <RAR project prefix> RAR Designation Cat. Set** or **Categories - <RAR project prefix> RAR Issue Cat. Set**.

When you expand the list, you see the available designation or issue categories and also a [Not Set] node which you can use to view documents that have not been categorized in your project.

![Field tree example](image)

Click on the desired field tag to view the corresponding documents.

### 3.7 Modifying issue importance

When you save a project Relativity automatically creates the Assisted Review Issue Importance RDO and populates it with the issue choices attached to the key issue field selected on the project. You can find this object and the issue choices at the bottom of the project settings layout.

By default, all issues have the same Medium Importance value. Any issue choices added to or removed from the key issue field are automatically reflected in this list.

If you want to make an issue more or less important than its default value of Medium, you can modify the relative importance of that issue. Doing this can provide more insight later when you refer to the Designation-Issue Comparison report.

To modify issue importance:

1. Click the **Sample-Based Learning** tab and select the project you wish to edit.
2. Scroll down to the **Assisted Review Issue Importance** list on the project settings page.
3. Click **Edit** next to a listed issue to change the level of importance.
4. In the **Importance** field, select one of the following weight values to apply to the issue:
   - High (500%)
   - Medium (100%)
   - Low (10%)

Changing the importance weighs the more important issues more heavily than the less important issues when calculating the total Issues Rank on the Designation-Issue Comparison report.

See [Designation-Issue Comparison report on page 89](#) for more information.

## 3.8 Deleting a project

To delete a project, you must have the Delete Object Dependencies permission, which is under Admin Operations. To confirm or to get this permission:

1. Navigate to the **Workspace Details** tab
2. Click the **Manage Workspace Permissions** link in the Relativity Utilities console.
3. Click **Edit Permissions**.
4. If the box next to **Delete Object Dependencies** under Admin Operations is unselected, select it.
5. Select the **Delete** radio button on all of these objects. You must also have the delete permission on all Assisted Review Objects.

To delete a project from the project layout:

1. Click **Delete**.
2. Click **OK** on the confirmation message. This message states that you are about to delete all rounds and audits associated with the project and that the categorization sets associated with the project won't be automatically deleted.

**Note:** A project with saved categorization results may take longer to delete than a project that has not been finalized. Allow extra time for a finalized project to be deleted.

To delete a project through a mass operation, perform the following:

1. Select the checkbox next to the project you want to delete.
2. Select **Delete** from the mass operations drop-down menu in the lower left corner of the view.
3. You will receive a pop-up stating that deleting this project will delete children and unlink associative objects. Click **Delete**.
4 Rounds

A round is a segment of a project designed to move a document set closer to a more relevant and defensible state. Most rounds are performed to train the system to recognize the difference between responsive and non-responsive documents, but every project also includes a round dedicated to quality control, specifically to check on the accuracy of Relativity’s coding decisions.

Note the following key Sample-Based Learning concepts:

- **Coded** – a document is referred to as coded or tagged when it has a choice set in the Designation field. Reviewers perform this part of Sample-Based Learning.

- **Categorized** – a document is referred to as categorized when it has a category set by Analytics. This is the automated part of review visible in the system-created fields after your first categorization.

Using rounds

Let's say that the Sample-Based Learning project you're managing started with 800 GB of data. You used keywords to get a set of three million documents down to around 160,000. You then indexed these documents and created your project. Your legal team conducted three good rounds of training, during which they reviewed a random sample of 1,000 documents each round. For each round, 7% of the documents were identified as responsive.

You feel that it's time to test Relativity's accuracy in what it determined was and wasn't responsive. You and your team conduct a fourth round for QC, during which you take a statistical sample of documents based on 95% confidence level and a 2% margin of error. You're aiming for a sample of under 2,500 documents with around 7% of those documents coded as responsive. After a single QC round, this is exactly what you get, telling you that your project is close to being done.

4.1 Preparation before starting Sample-Based Learning rounds

Before starting a round refer to the following:

- [Relativity Assisted Review on page 5](#) for an overview of Sample-Based Learning.
- [Sample-Based Learning on page 9](#)
- [Project setup on page 23](#) for information on project creation

4.2 Types of rounds

There are four types of rounds. Each one has a purpose and can occur throughout the process. The type of round works in tandem with the saved search selected.

- **Control set rounds** - For Relativity to automatically measure precision, recall, and determine the F1 score, it is necessary to set aside a control set of documents not utilized for training the system (not set as examples or excerpted). It is a best practice to create a control set as early as possible in the project, ideally as the very first round. See [Control set round on page 40](#).
- **Training rounds** - This is used to train the system. This is generally used initially prior to any other rounds, with the exception of the Control Set, if used. Training will occur until a significant portion of documents have a categorized value. See **Training round on page 50**.

- **Pre-coded seed rounds** - To utilize previously-coded documents, you can select pre-coded sets of documents. These documents cannot be part of a currently active control set round. See **Pre-coded seed round on page 45**.

- **Quality control (QC)** - This is sampling used to verify the categorization process. This will use one of the saved searches that includes documents with a value applied by the categorization process. Use quality control after you’ve already sufficiently trained the system. See **Quality Control (QC) round on page 56**.

### 4.3 Monitoring round progress

You can monitor round execution progress via the progress bar on the round layout. The progress bar displays the time elapsed and the percentage complete for the round.

![Progress bar for RAR Sample Set creation](image)

The progress bar on the project layout reflects when sample set creation is complete and the project is ready for review. Once the project displays a status of **RAR Sample Set created** you can begin review.

![Project status RAR Sample Set created](image)

From the project home page, you can click to launch a standalone status bar pop-up, which you can view while working outside your project or Relativity entirely.
- The standalone progress bar updates automatically.
- Only one standalone progress bar can be open per project.
- You receive an error if you are logged out of Relativity.

While review is in progress for the round, you can view that progress through the progress bar on the project home layout, which provides how many documents out of the sample set have been reviewed, as well as the corresponding percentage complete.

If you've enabled auto-batching on the round, you can view the progress of the batching for the sample set in its own bar.

You can view round details by clicking on the Round name on the project's home page.

Monitor the progress of the finishing round with the progress bar(s) on the project home layout.

The top bar reflects the overall status for all phases of round finish:
- Analyzing documents
- Saving documents (if selected when finishing the round)
- Categorizing documents
- Creating reports

The subsequent bars display the progress of each individual phase. These progress bars are determined by the options you selected when you initiated the round finish. For example, if you only chose to categorize designation, you would see the progress of only designation categorization.
4.4 Deleting a round

If you need to clear any mistakes made during round creation, you can delete a round for which no sample documents have been newly coded without removing any data from the project or workspace.

The option to delete a round is available underneath the Finish Round option on the console when a sample set has been created but no new documents have been coded for designation or issues. It is available under the Start Round option only when a round was canceled by the system.

To delete a round:
1. Click **Delete Round**.
2. Click **OK** on the confirmation. The confirmation informs you that deleting a round:
   - Permanently removes it from the project
   - Deletes the sample set saved search
   - Disassociates the sample set documents previously associated with the round
   - Removes the value set for the Use as example field on the sample documents

**Notes:**
- When you initiate the deletion of the round, all affected reports update in real time.
- A deleted round does not display in reports.
- Relativity removes a deleted round from the list of rounds.
- Round deletion is audited.
- The next round created after a round is deleted replaces the deleted round and has the same number.
- If you code or un-code the only coded document(s) in the round, the Delete Round option becomes available again.
- You can delete a round canceled by the system, but only the last round that was canceled.

### 4.5 Control set round

For Relativity to automatically measure the stability of the project (precision, recall, and F1 score), it is necessary to set aside a control set of documents that will not be utilized for training the system. **Control sets** (also known as truth sets) represent a random unbiased, statistically significant sample from the entire project universe. They can be thought of as a miniature model of an Assisted Review project.

It is recommended that you create a control set round as early as possible in the project (ideally as the very first round). The later you run one, the fewer eligible documents you will have. The documents contained within a control set sample are not used as examples, but they are categorized by reviewers. These specially treated documents allow an admin to track an Assisted Review project's status by observing how accurately the documents in the control set are categorized, which in return will offer insights into the accuracy of the entire project.

The **Control Set Statistics Report** is what is used to track a project's accuracy trends from round to round. The report displays precision, recall, and F1 scores for each round in a single line chart.

- Precision denotes the accuracy of the positive result documents (e.g., Responsive designation) which were categorized by the system.
- Recall indicates how many of the total positive result documents (e.g., Responsive designation) wereidentified by the system.
- The F1 score is the harmonic mean (a weighted average) of precision and recall.

Ideally, these scores should increase from round to round, and get to as close to 100% as is practical or possible for the project.
4.5.1 Executing a control set round

Notes:
- If you add new documents to the project, you invalidate the current control set. You must create a new one for Relativity to accurately determine precision and recall.
- If you have a large data set with low richness (e.g., very few responsive docs), you may need a larger control set.
- Documents categorized prior to being put in the control set can’t create overturns.
- Previously coded documents are not eligible to be added to a control set.

To execute a control set round:

1. Click **Start Round** on the console.
2. Select **Control set** as the **Round Type**.

![Round Information](image)

Note that if you create an **additional control set** later in the project:

- The new set replaces the previous one.
- Any documents from the inactive control set that are **coded** are eligible to be used in a Pre-coded seed round.
- Documents from the inactive control set that are **not coded** are eligible for a Training or QC round or they may be included in your new Control set round.
- Save a copy of the previously generated Control Set Statistics report before you start the new control set round if you wish to compare the two.

3. For the **Saved search for sampling**, it is recommended that you select the saved search used in the **Documents to be categorized** field on the project. This ensures that you get a random sample of documents from your entire project. It’s recommended that uncategorizable documents and unclustered documents not appear in this search.
4. Specify your desired **Sampling Methodology** settings.

**Note:** The fields in the Sampling Methodology section are defaulted to the values on the project settings.

- **Statistical sampling** - creates a sample set based on statistical sample calculations, which determines how many documents your reviewers need to code in order to get results that reflect the project universe as precisely as needed. Selecting this option makes the Margin of error field required.
  - **Confidence level** - the probability that the rate in the sample is a good measure of the rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.
  - **Margin of error** - the predicted difference between the observed rate in the sample and the true rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.

- **Percentage** - creates a sample set based on a specific percentage of documents from the project universe. Selecting this option makes the Sampling percentage field required.
  - Sampling percentage - the percentage of the eligible sample population used to create the sample size.

- **Fixed sample size** - creates a sample set based on a specific number of documents from the project universe. Selecting this option makes the second Fixed sample size field required.
  - **Fixed sample size** - the number of documents you want to include in your sample size.

5. Click **Calculate sample** to display a calculation of the sample based on the documents eligible for sampling and the values specified in the Sampling Methodology section. Clicking **Calculate sample** displays the number of documents in the saved search selected for the round and the number of documents in the sample. If the values for the sample and/or saved search are unexpected, you can change any setting in the Start Round layout and re-calculate before clicking **Go**. You can't calculate the sample if you don't have access to the saved search selected for the round.

![Sample Calculator](image)

**Sample Calculator**

**Calculate sample**

*Saved search: 2,492 documents - in All Documents*

*Sample size: 1,574 documents - based on 95% Confidence level and +/- 1.5% Margin of error and Round type of Training*

6. Specify how to batch documents out for review.

- **Automatically create batches** - determines whether or not a batch set and batches are automatically created for this round's sample set. By default, this field is set to whatever value was specified in the project settings. Once the sample set has been created, you can view and edit the corresponding batch set in the Batch Sets tab.
- **Maximum batch size** - the maximum number of documents that the automatically-created batches will contain. This is required if the Automatically create batches field above is set to Yes. This value must be greater than zero or an error appears when you attempt to save the project. The batch set and batches created from this project are editable after you create the project. By default, this field is set to whatever value was specified in the project settings.

7. Click **Go**.

8. Proceed to **Reviewing documents for a control set round below**.

### 4.5.2 Reviewing documents for a control set round

When reviewing documents for a control set round, you are only considering the responsiveness of the document, not setting documents as good examples. If you do not code documents included in the control set sample, those documents are eligible to be included in a subsequent control set round's sample set. See [Sample-Based Learning document review on page 20](#) for more information on protocol for assigning documents out and reviewing documents during a round.

**Note:** If you're done using a project, it's better for workspace performance if you finish the round rather than leaving it in a status of either **Review in Progress** or **Review complete**.

### 4.5.3 Finishing a control set round

Once all of the documents in the sample set have been coded, you should finish the round. You also have the option of finishing a round before all of the sample set documents have been coded.

**Note:** When you finish a control set round, if any documents are excerpted, marked as examples or uncoded, a warning appears on the Finish Round layout. If you continue to finish the round without manually removing the examples or excerpted text or coding the remaining documents, these documents are removed from the control set for reporting purposes.

If you need to find documents that were removed from a control set round, you can filter on or search for documents where the RAR Sample Set is the control set round name and Use as Example is set to True or the Designation excerpt field contains text. You can re-include the documents in the control set at a later time by switching the Use as Example field to No or removing the text excerpt.

Also, any documents in an active control set which are not coded are NOT eligible for sampling in subsequent rounds (except for a new control set round).

To finish a control set round:
1. Click **Finish Round** on the console.

![Assisted Review interface screenshot](image.png)

The Finish Round pop-up displays.

2. Click **Go** on the Finish Round pop-up to finish the round.

**Note:** If the control set is the first round of the project, reports aren't available when you finish that round; reports are only available after you finish your first non-control set round.

### 4.5.4 Reviewing Assisted Review reports after a control set round

If this round is done, it should be the first round.

The following reports should be reviewed after a control set round:
- **Control Set Statistics report** - provides the richness rate (total documents coded responsive) and the totals of how many coded for both responsive and not responsive documents. See Control Set Statistics report on page 80.

### 4.6 Pre-coded seed round

A pre-coded seed round only includes documents that have already been coded during manual review for the purpose of expediting training and are not included as part of another round. RAR identifies sample documents based on those found in the specified saved search with a value set on the Designation field.

**Notes:**
- Documents included in the pre-coded seed round sample MUST be reviewed using the proper Sample-Based Learning Protocol. See Sample-Based Learning standards and protocol on page 21.
- Ensure that the documents are coded using the same designation field as is set on the project. Also, ensure that you don't have an excessive amount (thousands) of documents in the sample when you start the round.

### 4.6.1 Executing a pre-coded seed round

**Notes:**
- It is important that a pre-coded seed round contain documents with both responsive and not responsive designations to give the system enough information to properly categorize documents in your project.
- The system does not consider the Use as an example field in determining eligibility. Therefore, if a document is coded on the designation field and is not currently in another round in the project, that document is eligible for inclusion in a pre-coded seed set even if the Use as example field is set to No or blank (not set).
- If the documents used in a Pre-coded seeds round are coded after the project is created, the Seed Reviewer field will be filled for any overturns caused by these pre-coded seeds. However, if the documents were coded via Mass Edit, the Seed Reviewer field will be empty.
- If the documents used in a Pre-coded seeds round were coded before the project was created, the Seed Reviewer field will be empty for any overturns caused by these pre-coded seeds.
- You can’t use pre-coded sets for quality control. If documents in the pre-coded seed rounds have already been categorized, they won’t create overturns.

To execute a pre-coded seed round:
1. Click **Start Round** on the console.
2. Select **Pre-coded seeds** as the **Round Type**.

![Round Information](image)

3. For the **Saved search for sampling**, select a search that contains all of the previously coded documents which you would like to include. If the search also includes documents that haven’t been coded, they are automatically excluded when the sample is created.

4. Specify your desired **Sampling Methodology** settings. The sample set is the randomly-selected group of documents produced by to be used for manual review as a means of training the system.

**Note:** The fields in the Sampling Methodology section are defaulted to the values on the project settings; however, if you select Training as the round type, you override those default values.

- **Statistical sampling** - creates a sample set based on statistical sample calculations, which determines how many documents your reviewers need to code in order to get results that reflect the project universe as precisely as needed. Selecting this option makes the Margin of error field required.
  - **Confidence level** - the probability that the rate in the sample is a good measure of the rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.
  - **Margin of error** - the predicted difference between the observed rate in the sample and the true rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.

- **Percentage** - creates a sample set based on a specific percentage of documents from the project universe. Selecting this option makes the Sampling percentage field required.
  - Sampling percentage - the percentage of the eligible sample population used to create the sample size.

- **Fixed sample size** - creates a sample set based on a specific number of documents from the project universe. Selecting this option makes the second Fixed sample size field required.
- **Fixed sample size** - the number of documents you want to include in your sample size.

5. Click **Calculate sample** to display a calculation of the sample based on the documents eligible for sampling and the values specified in the Sampling Methodology section. Clicking Calculate sample displays the number of documents in the saved search selected for the round and the number of documents in the sample. If the values for the sample and/or saved search are unexpected, you can change any setting in the Start Round layout and re-calculate before clicking **Go**. You can't calculate the sample if you don't have access to the saved search selected for the round. This button is disabled if you've selected Stratified sampling as the sampling type.

6. Specify how to batch documents out for review.
   - **Automatically create batches** - determines whether or not a batch set and batches are automatically created for this round's sample set. By default, this field is set to whatever value was specified in the project settings. Once the sample set has been created, you can view and edit the corresponding batch set in the Batch Sets tab.
   - **Maximum batch size** - the maximum number of documents that the automatically-created batches will contain. This is required if the Automatically create batches field above is set to **Yes**. This value must be greater than zero or an error appears when you attempt to save the project. The batch set and batches created from this project are editable after you create the project. By default, this field is set to whatever value was specified in the project settings.

7. Click **Go**.

8. Proceed to **Reviewing documents for a pre-coded seed round below**.

   **Note:** When the round is initially created, the field specified as the Use as an example field is set to **Yes** by default for documents included in the round. If you delete a round, RAR reverts the Use as an example field value to **Not Set** (null).

### 4.6.2 Reviewing documents for a pre-coded seed round

Documents for a pre-coded seed round do not need to be reviewed by a set of reviewers. See [Sample-Based Learning document review on page 20](#) for more information on protocol for assigning documents out and reviewing documents during a round.

   **Note:** If you're done using a project, it's better for workspace performance if you finish the round rather than leaving it in a status of either **Review in Progress** or **Review complete**.
4.6.3 Finishing a pre-coded seed round

Once all of the documents in the sample set have been coded, you should finish the round. You also have the option of finishing a round before all of the sample set documents have been coded.

To finish a pre-coded seed round:

1. Click **Finish Round** on the console.

   ![Console screenshot](image)

2. Specify whether you want to categorize documents when you finish the round. You have two options depending on your project:
   - **Categorize for designation** - allows you to categorize all documents in the project based on their designation coding.
   - **Categorize for issues** - allows you to categorize all documents in the project based on their issue coding. This is only available if you have added a key issue field to the project and a reviewer has issue-coded at least one document in the sample set.

3. Specify whether you want to save categorization results from the previous round when you finish the current round. You may have two options depending on your project:
- **Save designation results** - allows you to save the results of designation coding from the previous categorization. This is useful because when categorization runs, the previous results are cleared in order to apply the new category values. You can't save designation results if you did not categorize designations in a previous round.

- **Save issue results** - allows you to save the results of issue coding from the previous categorization. This is only available if you have added a key issue field to the project. You can only save issue results if you categorized issues in a previous round.

**Note:** You shouldn't save results at the end of every round. Saving results, especially for larger cases, can add several hours to the time it takes to finish the round.

4. Enter the naming for your categorization results.

- **Categorization results set name** - the name of the categorization results set. By default, this is the name of the previous round. This is only available for editing if you are saving designation and/or issue results.

- **Categorization results set description** - a description of the categorization results. This is only available for editing if you are saving designation and/or issues results.
5. Click Go. If you choose to both categorize and save results, the saving of results is performed first, then categorization.

4.6.4 Reviewing Sample-Based Learning reports after a pre-coded seed round

The following reports should be reviewed after a pre-coded seed round:

- **Round Summary report** – useful after categorization because it shows the changes in categorization percentage from round to round and also provides categorization volatility. See [Round Summary report on page 65](#).

**Note:** If issues are also being categorized by Sample-Based Learning, you can also review the [Issue Reports](#).

4.7 Training round

A training round is intended to teach Sample-Based Learning how to interpret more uncategorized documents so that it is more likely to successfully categorize others with a high level of confidence and accuracy.

4.7.1 Executing a training round

To execute a training round:

1. Click **Start Round** on the console.
2. Select **Training** as the **Round Type**. When you select this as the round type the Sampling type field defaults to Stratified sampling.

3. Enter a **Round description**.
4. For the **Saved search for sampling**, select the automatically created `<Project Saved Search>` search. This is the only search available and it contains only uncategorized documents because you haven’t started the project at this point. See [Viewing categorized and uncategorized document saved searches on page 30](#) for more information. After the initial training round has been com-
completed, and categorization has run, subsequent training round saved searches should typically be set to uncategorized documents.

5. Specify your desired **Sampling Methodology** settings. The sample set is the randomly-selected group of documents produced by to be used for manual review as a means of training the system. Stratified sampling is selected by default.

**Note:** The fields in the Sampling Methodology section are defaulted to the values on the project settings; however, if you select Training as the round type, you override those default values.

- **Stratified sampling** - groups the round saved search documents into subgroups based on the documents’ concepts and returns the documents that cover the vast majority of the conceptual space or until the **Maximum sample size** or **Minimum seed influence** has been met. This type of sampling allows RAR to effectively train with as few documents as possible. Selecting this type makes the Maximum sample size and Minimum seed influence fields available and disables the Calculate sample button. The Stratified sampling option is only available when you select the **Training** round type.

**Notes:**
- You can increase your categorization coverage of your conceptual space by running multiple stratified rounds.
- If all documents in the sample were used as examples, they would categorize 80% of the documents
- You can limit the sample by **Maximum sample size** or **Minimum seed influence**.
- If you decrease the coherence value on the Designation Categorization Set associated with your RAR project, RAR will return fewer documents for a stratified sample and categorize more documents with each example.
- You should still follow best practices for excerpting and selecting documents to use as examples when reviewing documents from a stratified sample. See the Analytics Guide for more information.

- **Maximum sample size** - the maximum number of documents you want returned in a stratified sample. For example, if you set the value at 500, the round will contain the 500 strongest documents (according to their seed influence). The stratified sample is created once this or the Minimum seed influence value is met. Leaving this field blank means that the round will include all documents returned by the stratified sample, unless you've entered a Minimum seed influence.

- **Minimum seed influence** - the minimum number of documents required to be categorized by each example document returned in a stratified sample. For example, if you leave this at its default value of 25, every document returned in the sample will categorize at least 25 other documents if they are designated as examples. This field is only available for training rounds with a stratified sampling type. The stratified sample is created once this or the Maximum sample size value is met.

- **Statistical sampling** - creates a sample set based on statistical sample calculations, which determines how many documents your reviewers need to code in order to get results that reflect the project universe as precisely as needed. Selecting this option makes the Margin of
error field required.

- **Confidence level** - the probability that the rate in the sample is a good measure of the rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.

- **Margin of error** - the predicted difference between the observed rate in the sample and the true rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.

- **Percentage** - creates a sample set based on a specific percentage of documents from the project universe. Selecting this option makes the Sampling percentage field required.
  - Sampling percentage - the percentage of the eligible sample population used to create the sample size.

- **Fixed sample size** - creates a sample set based on a specific number of documents from the project universe. Selecting this option makes the second Fixed sample size field from the project universe. Selecting this option makes the second Fixed sample size field required.
  - Fixed sample size - the number of documents you want to include in your sample size.

6. Click **Calculate sample** to display a calculation of the sample based on the documents eligible for sampling and the values specified in the Sampling Methodology section. Clicking **Calculate sample** displays the number of documents in the saved search selected for the round and the number of documents in the sample. If the values for the sample and/or saved search are unexpected, you can change any setting in the Start Round layout and re-calculate before clicking Go. You can't calculate the sample if you don't have access to the saved search selected for the round. This button is disabled if you've selected Stratified sampling as the sampling type.

7. Specify how to batch documents out for review.

- **Automatically create batches** - determines whether or not a batch set and batches are automatically created for this round's sample set. By default, this field is set to whatever value was specified in the project settings. Once the sample set has been created, you can view and edit the corresponding batch set in the Batch Sets tab.

- **Maximum batch size** - the maximum number of documents that the automatically-created batches will contain. This is required if the Automatically create batches field above is set to Yes. This value must be greater than zero or an error appears when you attempt to save the project. The batch set and batches created from this project are editable after you create the project. By default, this field is set to whatever value was specified in the project settings.

8. Click **Go**.
9. Proceed to Training round document review below.

**Note:** When the round is created, the field specified as the Use as an example field is set to Yes by default for documents included in the round. If you delete a round, Sample-Based Learning reverts the Use as an example field value to Not Set (null).

### 4.7.2 Training round document review

Sample-Based Learning is trained as documents are reviewed and assigned to categories. See Sample-Based Learning document review on page 20 for more information on protocol for assigning documents out and reviewing documents during a round.

**Note:** If you’re done using a project, it's better for workspace performance if you finish the round rather than leaving it in a status of either Review in Progress or Review complete.

### 4.7.3 Finishing a training round

Once all of the documents in the sample set have been coded, you should finish the round. You also have the option of finishing a round before all of the sample set documents have been coded.

To finish a training round:
1. Click **Finish Round** on the console.

2. Specify whether you want to categorize documents when you finish the round. You have two options depending on your project:
   - **Categorize for designation** - allows you to categorize all documents in the project based on their designation coding.
   - **Categorize for issues** - allows you to categorize all documents in the project based on their issue coding. This is only available if you have added a key issue field to the project and a reviewer has issue-coded at least one document in the sample set.

3. Specify whether you want to save categorization results from the previous round when you finish the current round. You may have two options depending on your project:
   - **Save designation results** - allows you to save the results of designation coding from the previous categorization. This is useful because when categorization runs, the previous results
are cleared in order to apply the new category values. You can’t save designation results if you did not categorize designations in a previous round.

- **Save issue results** - allows you to save the results of issue coding from the previous categorization. This is only available if you have added a key issue field to the project. You can only save issue results if you categorized issues in a previous round.

**Note:** You shouldn’t save results at the end of every round. Saving results, especially for larger cases, can add several hours to the time it takes to finish the round.

4. Enter the naming for your categorization results.

   - **Categorization results set name** - the name of the categorization results set. By default, this is the name of the previous round. This is only available for editing if you are saving designation and/or issue results.

   - **Categorization results set description** - a description of the categorization results. This is only available for editing if you are saving designation and/or issues results.

5. Click **Go**. If you choose to both categorize and save results, the saving of results is performed first, then categorization.
4.7.4 Reviewing Sample-Based Learning reports after a training round

The following reports should be reviewed after finishing a training round:

- **Round Summary Report** – useful after categorization because it shows the changes in categorization percentage from round to round. See Round Summary report on page 65.
- **Control Set Statistics** – tracks progress of precision and recall and F1. See Control Set Statistics report on page 80
- **Rank Distribution** – shows level of conceptual similarity between human-coded documents and the overall categorized documents. See Rank Distribution report on page 70
- **Project Summary** – tracks overall project health. You can see a snapshot of overturn and categorization results as well as control set statistics in one place. See Project Summary report on page 76

**Note:** If issues are also being categorized by Assisted Review, you can also review the Issue Reports.

4.8 Quality Control (QC) round

A quality control round is intended to provide reviewers with documents that Sample-Based Learning categorized so that it can calculate an overturn rate by comparing the categorized value with the value manually coded during this round.

4.8.1 Executing a QC round

To execute a quality control round:

1. Click **Start Round** on the console.
2. Select **Quality control** as the **Round Type**.

3. For the **Saved search for sampling**, select a saved search containing categorized documents (e.g., `<Project Saved Search> - Categorized`). See Viewing categorized and uncategorized document saved searches on page 30 for more information.
4. Specify your desired **Sampling Methodology** settings. The sample set is the randomly-selected group of documents produced by to be used for manual review as a means of training the system.

**Note:** The fields in the Sampling Methodology section are defaulted to the values on the project settings; however, if you select Training as the round type, you override those default values.

- **Statistical sampling** - creates a sample set based on statistical sample calculations, which determines how many documents your reviewers need to code in order to get results that reflect the project universe as precisely as needed. Selecting this option makes the Margin of error field required.
  - **Confidence level** - the probability that the rate in the sample is a good measure of the rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.
  - **Margin of error** - the predicted difference between the observed rate in the sample and the true rate in the project universe. This is used in the round to calculate the overturn range as well as the sample size, if you use statistical sampling.

- **Percentage** - creates a sample set based on a specific percentage of documents from the project universe. Selecting this option makes the Sampling percentage field required.
  - Sampling percentage - the percentage of the eligible sample population used to create the sample size.

- **Fixed sample size** - creates a sample set based on a specific number of documents from the project universe. Selecting this option makes the second Fixed sample size field required.
  - **Fixed sample size** - the number of documents you want to include in your sample size.

5. Click **Calculate sample** to display a calculation of the sample based on the documents eligible for sampling and the values specified in the Sampling Methodology section. Clicking **Calculate sample** displays the number of documents in the saved search selected for the round and the number of documents in the sample. If the values for the sample and/or saved search are unexpected, you can change any setting in the Start Round layout and re-calculate before clicking Go. You can't calculate the sample if you don't have access to the saved search selected for the round. This button is disabled if you've selected Stratified sampling as the sampling type.

![Sample Calculator](image)

- **Sample Calculator**
  - **Calculate sample**

  **Saved search: 2,492 documents - in All Documents**
  **Sample size: 1,574 documents - based on 95% Confidence level and +/- 1.5% Margin of error and Round type of Training**

6. Specify how to batch documents out for review.

- **Automatically create batches** - determines whether or not a batch set and batches are automatically created for this round's sample set. By default, this field is set to whatever value
was specified in the project settings. Once the sample set has been created, you can view and edit the corresponding batch set in the Batch Sets tab.

- **Maximum batch size** - the maximum number of documents that the automatically-created batches will contain. This is required if the Automatically create batches field above is set to Yes. This value must be greater than zero or an error appears when you attempt to save the project. The batch set and batches created from this project are editable after you create the project. By default, this field is set to whatever value was specified in the project settings.

7. Click **Go**.

8. Proceed to **Reviewing documents for a QC round below**.

**Note:** When the round is created, the field specified as the Use as an example field is set to Yes by default for documents included in the round. If you delete a round, RAR reverts the Use as an example field value to Not Set (null).

### 4.8.2 Reviewing documents for a QC round

When reviewing documents for a QC round, you only review categorized documents. You are testing the accuracy of the categorized results of your Sample-Based Learning project. Reviewers are coding documents that have already been categorized by the system. They will not be aware of this fact. They will code as normal, but the system will evaluate their coding values against documents that the system categorized. See [Sample-Based Learning document review on page 20](#) for more information on protocol for assigning documents out and reviewing documents during a round.

**Note:** If you’re done using a project, it’s better for workspace performance if you finish the round rather than leaving it in a status of either **Review in Progress** or **Review complete**.

### 4.8.3 Reviewing Sample-Based Learning reports during a QC round

After the QC round is started and all the documents in the sample are coded, admins assign the seed documents out after review after reading the reports to be corrected. Reviewers make the corrections to any seed documents that are causing issues (see [Evaluating overturns and making corrections on the next page](#)).

The following reports should be reviewed after QC round sample documents have been coded but before finishing the round:

- **Round Summary report** – useful after categorization because it shows the changes in categorization percentage from round to round. Also provides categorization volatility. See [Round Summary report on page 65](#).

- **Control Set Statistics report** – tracks progress of precision and recall and F1. Also gives the Summary of Rounds. See [Control Set Statistics report on page 80](#).

- **Overturn summary report** – tracks overturn percentages round to round. There are no overturns prior to a QC round. See [Overturn Summary report on page 71](#).

- **Viewing overturned documents** - The Overturned Documents view allows an Assisted Review admin to view documents that require re-evaluation quickly and efficiently. You can focus on a single round and filter by the highest ranking overturns. You may also use the pivot feature to see

![Relativity](https://example.com/relativity.png)

Assisted Review Sample-Based Learning Guide 58
the most influential seed documents (documents that are responsible for a large number of overturns). Once you identify documents as needing further analysis, you can click on a link in order to review the document immediately. See Viewing overturned documents on page 95.

- **Rank Distribution report** – shows level of conceptual similarity between human coded documents and the overall categorized documents. See Rank Distribution report on page 70.
- **Project Summary report** – tracks overall project health. You can see a snapshot of overturn and categorization results as well as control set statistics in one place. See Project Summary report on page 76.

**Note:** If issues are also being categorized by Assisted Review, you can also review the Issue Reports.

### 4.8.4 Evaluating overturns and making corrections

During a computer-assisted review, a case team moves through several rounds of coding to train the system on the document collection and validate the computer’s results. This isn’t a formal round, but a between-rounds workflow used to make any necessary adjustments to the project to prepare for the next round. It consists of identifying, analyzing, and correcting (re-coding) documents which have a significant and adverse effect on project results. You are finding the seed documents that caused the overturns and then making any coding corrections to those seed documents that need to be made.

Potential coding errors are reported in the Overturned Documents link in a Relativity project’s console. The Overturned Documents view allows an Assisted Review admin to view documents that require re-evaluation quickly and efficiently. You can focus on a single round and filter by the highest ranking overturns. You may also use the pivot feature to see the most influential seed documents (documents that are responsible for a large number of overturns). Once you identify documents as needing further analysis, you can click on a link in order to review the document immediately. See Viewing overturned documents on page 95.

**Note:** We recommend that you make these adjustments prior to finishing a round and categorizing documents. This allows the system to make use of the corrections performed, and then apply them to the next true round.

Consider the following when your reviewers are evaluating overturns and making corrections:

**Correcting coding inconsistencies between true or conceptual duplicates:**

- Each seed-overturn pair has a rank (or score) which indicates the degree of conceptual similarity they share. The maximum possible score is 100, which means the two documents are conceptual duplicates. Conceptual duplicates are documents which may or may not have identical text, but do contain the same conceptual content according to the analytics index. While it is possible that conceptual duplicates may also be exact textual duplicates (i.e., documents with the same MD5 hash value), this should not be assumed from a score of 100.

- We recommend that you use the Overturn Documents report to locate these documents by filtering on the round and sorting by descending rank. A good best practice is to re-evaluate each seed-overturn pair having a rank of 95 and higher to see which document was coded correctly, as well as whether each is a suitable example.

**Identifying and correcting the most influential seed documents:**
When viewing overturn reports at the end of a round, the same few documents can be responsible for many overturns. If those seed documents were incorrectly coded, they can greatly inflate the overturn rate for the entire project. Finding and correcting these situations is an essential component of QC round protocol.

The quickest way to find the most influential documents is by using Pivot on the Overturned Documents report. Simply choose Seed document in the Group by drop-down and leave the <Total Only> drop-down as is.

**Using the Overtur Analysis related items pane:**

- Once a document has been targeted for re-evaluation during a QC round, you can navigate directly to it using the hyperlinks in the Overturned Documents report. Once you reach the core reviewer interface, open the Overtur Analysis related items pane by clicking the RAR icon in the bottom right corner.
- Clicking the file icon next to the document's control number opens the target document in a separate viewer window. This allows a reviewer to compare the two documents side by side to assist in the decision-making process.

**Note:** The Overturned Documents view is helpful for review management, but you may want to prevent users from seeing the rest of the project when they only need access to overturns. You can also provide reviewers access to overturns via the field tree, which includes an overturn status field for your project. To pursue this option, create a view that can be used in conjunction with the field tree.

### 4.8.5 Finishing a QC round

Once all of the documents in the sample set have been coded, you should finish the round. You also have the option of finishing a round before all of the sample set documents have been coded.

To finish a QC round:
1. Click **Finish Round** on the console.

2. Specify whether you want to categorize documents when you finish the round. You have two options depending on your project:
   - **Categorize for designation** - allows you to categorize all documents in the project based on their designation coding.
   - **Categorize for issues** - allows you to categorize all documents in the project based on their issue coding. This is only available if you have added a key issue field to the project and a reviewer has issue-coded at least one document in the sample set.

3. Specify whether you want to save categorization results from the previous round when you finish the current round. You may have two options depending on your project:
   - **Save designation results** - allows you to save the results of designation coding from the previous categorization. This is useful because when categorization runs, the previous results are cleared in order to apply the new category values. You can't save designation results if you did not categorize designations in a previous round.
- **Save issue results** - allows you to save the results of issue coding from the previous categorization. This is only available if you have added a key issue field to the project. You can only save issue results if you categorized issues in a previous round.

**Note:** You shouldn't save results at the end of every round. Saving results, especially for larger cases, can add several hours to the time it takes to finish the round.

4. Enter the naming for your categorization results.
   - **Categorization results set name** - the name of the categorization results set. By default, this is the name of the previous round. This is only available for editing if you are saving designation and/or issue results.
   - **Categorization results set description** - a description of the categorization results. This is only available for editing if you are saving designation and/or issues results.

![Finish Round P03 003](image)

5. Click **Go**. If you choose to both categorize and save results, the saving of results is performed first, then categorization.
5 Sample-Based Learning reports

Sample-Based Learning’s reporting structure provides a detailed picture of how the project is progressing. In addition to gauging your progress, you can gain transparency into why Relativity categorized documents the way it did.

**Note:** This page uses Responsive and Not Responsive as designation choice terms because those are the most common. The designation choice terms you use in your project may be different.

**Using reports**

Imagine you’re the senior attorney on a case using Sample-Based Learning, and you’ve already reviewed a set of 1,500 randomly selected documents to train the system on what was and wasn’t responsive. The system then categorized the rest of the documents, applying the coding decisions you made to the conceptually similar documents.

After this first training round, you conduct four QC rounds to test Relativity’s accuracy in grouping the documents. During the QC rounds, another reviewer codes a random sample of documents without knowing if Relativity has categorized them as responsive or non-responsive. As the QC rounds proceed, you run the overturn summary report and notice an unusual number of documents being overturned by reviewers.

These overturns consisted largely of company newsletters that were showing up as responsive due only to their inclusion of relevant property and project names. This tells you that you can easily use keyword searching to remove the newsletters from the document universe. You can then rebuild the Analytics index to exclude these newsletters, thus saving you more time and eliminating useless information from your project.

5.1 Report types

5.1.1 Designation reports

Designation-related reports include the following options:

- **Round Summary report** - The Round Summary report is split into two graphs. The first graph provides a categorization breakdown of each category per round, allowing you to track categorization results as the Sample-Based Learning project progresses. The second graph displays the volatility of each category per round, while accompanying tables provide additional information such as document tallies and population size changes. See [Round Summary report on page 65](#).

- **Rank Distribution report** - The Designation Rank Distribution report offers insight into categorization results according to their rank score. The accompanying table also provides document tallies. See [Rank Distribution report on page 70](#).

- **Overtur Summary report** - The Overtur Summary report provides a high level view of a Sample-Based Learning project’s overturn percentages and tallies from round to round. The accompanying table gives even greater detail, including a breakdown of responsive and non-responsive overturn rates. See [Overtur Summary report on page 71](#).
5.1.2 Control set reports
Control set-related reports include the following options:
- **Control set statistics report** - The Control Set Statistics report displays a project's accuracy trends from round to round. The report displays precision, recall, and F1 scores for each round in a single line chart. See [Control Set Statistics report on page 80](#).

5.1.3 Issue reports
Issue-related reports include the following options:
- **Issue summary report** - The issue summary provides a breakdown of issue coding categorization results. The table version of the report also displays document tallies and previous round information. See [Issue Summary report on page 87](#).
- **Designation-Issue comparison report** - The Designation-Issue Comparison Report provides valuable insights into designation and issue coding categorization results, and how they relate to one another. See [Designation-Issue Comparison report on page 89](#).

5.1.4 RAR project document reporting views
- **Viewing control set documents** - See the [Viewing control set documents on page 94](#) section in [RAR project document reporting views on page 94](#).
- **Viewing overturned documents** - The Overturned Documents view allows a Sample-Based Learning admin to view documents that require re-evaluation quickly and efficiently. You can focus on a single round and filter by the highest ranking overturns. You may also use the pivot feature to see the most influential seed documents (documents that are responsible for a large number of overturns). Once you identify documents as needing further analysis, you can click on a link in order to review the document immediately. See [Viewing overturned documents on page 95](#).
- **Viewing saved categorization results** - The Saved Cat. Results view provides details on each set of saved categorization results based on what you selected when you finished a round. See [Viewing saved categorization results on page 98](#).

5.2 Viewing reports
You can access each report by clicking on its link on the Assisted Review project console.
The following toolbar appears at the top of each report:
This toolbar provides the following options, from left to right:

- Print opens a print dialog to select printing options.
- Export and save/file type allows you to save the report in the selected file type format (pdf, xls, xlsx, csv, or image).
- Each report contains the Close button at the top of the window. Click this to close the report.

**Note:** The choice you selected in the **Positive choice for designation** field on the project appears first in the graphs and tables of all reports containing designation choices. For example, if you selected Responsive as the positive choice, it appears before all other designation choices.

### 5.3 Designation reports

The Designation Reports section on the Assisted Review project console lists a variety of options for gathering insights into how documents were categorized for your Assisted Review project by round.

#### 5.3.1 Round Summary report

To access this report, click **View Round Summary** under **Designation Reports** on the console.

The **Round Summary** report provides the categorization results of the entire project after each round, as well as a comparison of the results from round to round. To access this report, you must have completed at least one round in the project.

**Note:** All numbers and percentage values in Assisted Review reports are based on the number of documents in the project at the end of the round.
The Designation Categorization Results graph provides the following data:

- **% of docs in the project** - the graph’s vertical axis; this contains percentages that correspond to the designation value bars.

- **Round** - the graph’s horizontal axis; this contains the name of each completed round in the project and a bar that provide a percentage of what documents in the project were categorized with what designation values (e.g., Responsive, Non Responsive, and Uncategorized).

The Designation Categorization Results tables provide the following data:
<table>
<thead>
<tr>
<th>Round name</th>
<th>Categorized Responsive</th>
<th>Categorized Non Responsive</th>
<th>Uncategorized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
</tr>
<tr>
<td>DRD-A 001</td>
<td>78</td>
<td>6.34%</td>
<td>232</td>
</tr>
<tr>
<td>DRD-A 002</td>
<td>81</td>
<td>6.58%</td>
<td>280</td>
</tr>
<tr>
<td>DRD-A 003</td>
<td>81</td>
<td>6.58%</td>
<td>280</td>
</tr>
<tr>
<td>DRD-A 005</td>
<td>81</td>
<td>6.58%</td>
<td>280</td>
</tr>
<tr>
<td>DRD-A 006</td>
<td>141</td>
<td>10.70%</td>
<td>346</td>
</tr>
<tr>
<td>DRD-A 007</td>
<td>84</td>
<td>31.34%</td>
<td>107</td>
</tr>
<tr>
<td>DRD-A 008</td>
<td>1,734</td>
<td>22.02%</td>
<td>6,119</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round Name</th>
<th>Manually Coded Documents (#)</th>
<th>Only Categorized Documents (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Responsive</td>
<td>Non Responsive</td>
</tr>
<tr>
<td>DRD-A 001</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>DRD-A 002</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>DRD-A 003</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DRD-A 005</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>DRD-A 006</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>DRD-A 007</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>DRD-A 008</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

- **Round name** - the name of the round, as generated automatically when the round was started
- **Categorized Responsive #** and % - the total number and percentage of documents in the project categorized as responsive
- **Categorized Non Responsive #** and % - the total number and percentage of documents in the project categorized as non responsive
- **Uncategorized #** and % - the total number and percentage of documents in the project that were not categorized
- **Manually Coded Documents (#)** - provides a count of documents that were manually coded as either responsive or non responsive for each round. This includes both documents that are examples and those that are not examples.
- **Only Categorized Documents (%)** - provides the percentage of categorized documents that fall into the responsive and non responsive categories

The **Categorization Volatility** graph and table provide the following data:
The **Categorization Volatility** table provides the following data:

- **Percent Volatility** - the graph’s vertical axis; this contains the percent change in categorization results from round to round. For example, between rounds 8 and 9, there was a 6% increase in Responsive documents.

- **Round** - the graph’s horizontal axis; this contains the name of each completed round in the project; the graph itself connects the percentage in each round so that you can view the change in percentage of documents coded with each designation value (Responsive and Not Responsive).
<table>
<thead>
<tr>
<th>Round name</th>
<th>Not Responsive (^\rightarrow) Responsive</th>
<th>Responsive (\rightarrow) Not Responsive</th>
<th>Uncategorized change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
</tr>
<tr>
<td>D86 001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D86 002</td>
<td>257</td>
<td>7.34%</td>
<td>737</td>
</tr>
<tr>
<td>D86 003</td>
<td>36</td>
<td>1.03%</td>
<td>244</td>
</tr>
<tr>
<td>D86 004</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>D86 005</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>D86 006</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>D86 007</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>D86 008</td>
<td>0</td>
<td>0.34%</td>
<td>0</td>
</tr>
<tr>
<td>D86 009</td>
<td>-47</td>
<td>6.38%</td>
<td>-531</td>
</tr>
</tbody>
</table>

- **Round name** - the name of the round, as generated automatically when the round was started.
- **Not Responsive \(\rightarrow\) Responsive** # and % - the total change in number of documents marked responsive for that round and the percentage of change plus or minus of documents marked as responsive since the last round, respectively.
- **Responsive \(\rightarrow\) Not Responsive** # and % - the total change in number of documents marked not responsive for that round and the percentage of change of documents marked as non responsive since the last round, respectively.
- **Uncategorized change** # and % - the total change in number of documents that were not categorized in that round and the percentage of change of documents that were not categorized since the last round, respectively.

**Note:** The example data in the Categorization Volatility graph and table represents a case in which documents are removed as examples during a meta round. Once the documents are no longer examples, it's possible that fewer documents would categorize.

The Summary of Rounds table provides the following data:

- **Round name** - the name of the round, as generated automatically when the round was started.
- **Round type** - the type of round specified when the round was started: training, quality control, pre-coded seeds, or control set.
- **Reviewed docs in round (#)** - the total number of documents reviewed in the round.
- **Reviewed docs in project (#)** - the total number of documents reviewed in the project.
- **Docs in project (#)** - the total number of documents, reviewed and unreviewed, in the project.
- **Round description** - the description of the round, as provided by the project manager when starting the round.
5.3.2 Rank Distribution report

The Rank Distribution report provides information on the level of conceptual similarity of the categorized documents to each documents’ seed, which is the human-reviewed document used as the basis for the system decision. This information is expressed with the percentage and number of documents categorized in rank ranges for each designation choice.

To access this report, click View Rank Distribution under Designation Reports on the console.

The Designation Rank Distribution graph provides the following data:

- % of categorized docs - the graph's vertical axis; it contains the percentage of categorized documents in the round.
- Designation rank - the graph's horizontal axis; it contains the contains ranges of values for the designation rank.
The Designation Rank Distribution table provides the following data:

<table>
<thead>
<tr>
<th>Designation rank</th>
<th>Responsive</th>
<th></th>
<th>Not Responsive</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
</tr>
<tr>
<td>100</td>
<td>7</td>
<td>1.01%</td>
<td>225</td>
<td>32.33%</td>
</tr>
<tr>
<td>95-99</td>
<td>0</td>
<td>0.00%</td>
<td>124</td>
<td>17.82%</td>
</tr>
<tr>
<td>90-94</td>
<td>0</td>
<td>0.00%</td>
<td>13</td>
<td>1.87%</td>
</tr>
<tr>
<td>85-89</td>
<td>3</td>
<td>0.43%</td>
<td>33</td>
<td>4.74%</td>
</tr>
<tr>
<td>80-84</td>
<td>26</td>
<td>3.74%</td>
<td>32</td>
<td>4.60%</td>
</tr>
<tr>
<td>75-79</td>
<td>75</td>
<td>10.78%</td>
<td>7</td>
<td>1.01%</td>
</tr>
<tr>
<td>70-74</td>
<td>135</td>
<td>19.46%</td>
<td>16</td>
<td>2.30%</td>
</tr>
</tbody>
</table>

- **Designation rank** - range of rank values in which a document falls for designation. The higher the rank, the more conceptually related the categorized document is to the closest seed document.

- **Responsive (#) and (%)** - number of documents categorized in the given Responsive Designation rank range and the percentage of total categorized documents.

- **Not Responsive (#) and (%)** - number of documents categorized in the given Not Responsive Designation rank range and the percentage of total categorized documents.

**5.3.3 Overturn Summary report**

The Overturn Summary report provides information on the number of overturns and estimated overturn range per round. An overturn occurs when a reviewer's manual coding value conflicts with the Assisted Review-assigned category. This report updates dynamically while a round is in progress.

To access this report, click View Overturn Summary under Designation Reports on the console.

The Overturns and Estimated Overturn Range graph provides the following data:
The Overturn Summary table gives a breakdown of overturns made per round and provides the following data:

- **% of docs in round** - the graph's vertical axis; this contains the percentage of documents in the round.
- **Round** - the graph's horizontal axis; this contains the name of each round in the project.
- **Overturns** - the blue dot; this represents the percent of eligible documents that were overturned in that round.
- **Estimated overturn range** - the light blue bar; this represents the estimated range of overturned documents in that round if you had manually reviewed all documents, not just the documents in the sample set.
The Overturns by Designation Category table gives a breakdown of the overturns per designation field choice and provides the following data:

<table>
<thead>
<tr>
<th>Round name</th>
<th>Overturned docs #</th>
<th>Overturned docs %</th>
<th>Docs eligible for overturn (#)</th>
<th>Estimated overturn range #</th>
<th>Estimated overturn range %</th>
<th>Calculated margin of error</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAR 001</td>
<td>7</td>
<td>29.17%</td>
<td>24</td>
<td>3,715 - 13,781</td>
<td>12.39% - 45.95%</td>
<td>16.78%</td>
</tr>
<tr>
<td>RAR 002</td>
<td>62</td>
<td>23.40%</td>
<td>265</td>
<td>2,678 - 4,139</td>
<td>18.39% - 28.41%</td>
<td>5.01%</td>
</tr>
<tr>
<td>RAR 003</td>
<td>64</td>
<td>23.97%</td>
<td>267</td>
<td>3,164 - 4,829</td>
<td>18.98% - 28.96%</td>
<td>4.99%</td>
</tr>
</tbody>
</table>

- **Round name** - the name of the round.
- **Overturned docs** - the number and percentage of documents that were manually coded with a different designation choice compared to the system categorized value. For example, a document was categorized by the system as Not Responsive and later coded manually as Responsive.
- **Docs eligible for overturn (#)** - the number of documents that are candidates to be overturned. This includes documents that have been categorized and are manually coded as part of the round's sample set. If 0 documents are eligible for overturn it means the round contains no documents which have already been categorized.
- **Estimated overturn range** - the statistically-based estimation of the range of documents for the entire round that would be overturned.
- **Calculated margin of error** - the percentage of random sampling error in the round's results. This is calculated based on the confidence level you select and the number of documents which have been manually reviewed.

The Overturns by Designation Category table gives a breakdown of the overturns per designation field choice and provides the following data:

<table>
<thead>
<tr>
<th>Round name</th>
<th>Responsive overturned docs #</th>
<th>Responsive docs eligible for overturn (#)</th>
<th>Not Responsive overturned docs #</th>
<th>Not Responsive docs eligible for overturn (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAR 001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RAR 002</td>
<td>7</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RAR 003</td>
<td>33</td>
<td>85</td>
<td>29</td>
<td>180</td>
</tr>
<tr>
<td>RAR 004</td>
<td>49</td>
<td>77</td>
<td>15</td>
<td>190</td>
</tr>
</tbody>
</table>

- **Round name** - the name of the round.
- **Responsive overturned docs #** and % - the number and percentage of documents previously categorized by the system as Responsive and overturned by being manually coded as Not Responsive in the current round. For example in Round 2, 7 documents categorized as Responsive were manually coded as Not responsive, and this number is 41.18% of the 17 Responsive docs eligible for overturn.
- **Responsive docs eligible for overturn (#)** - the number of documents categorized by the system and later coded manually as Responsive.

- **Not Responsive overturned docs #** and % - the number and percentage of documents previously categorized by the system as Not Responsive and overturned by being manually coded as Responsive in the current round. For example, in Round 3, 29 documents categorized as Not Responsive were manually coded as Responsive, and this number is 16.11% of the 180 Not Responsive docs eligible for overturn.

- **Not Responsive docs eligible for overturn #** - the number of documents categorized by the system and later coded manually as Not Responsive.

The **Summary of Rounds** gives a breakdown of all settings made per round and provides the following data:

- **Round name** - the name of the round.
- **Round type** - the type of round specified when the round was started: training, quality control, pre-coded seeds, or control set.
- **Confidence level** - the value set in the confidence level field when the round was started.
- **Reviewed docs (#)** - the number of documents reviewed in the round.
- **Reviewed docs in project (#)** - the total number of documents that have been reviewed in the project so far; this is a sum of all the Reviewed docs (#) values.
- **Docs in round (#)** - the number of documents returned in the saved search for that round.
- **Round description** - the description entered when the round was started.

**Note:** Uncategorized documents are not included in the Overturn Summary report calculations. This is why some rounds have a blank value, as there are no documents eligible for overturn.

### 5.3.3.1 Overturn Summary report formulas
RAR uses several formulas to arrive at the data found in the Overturn Summary report.

These formulas are based on a project with the following parameters:

- Documents to be categorized (project saved search) contains 50,000 documents
- Reviewed count is 2,000
- Saved search for sampling (round saved search) is 50,000
- Sampling methodology is:
  - Desired confidence level – 90%
  - Sampling type – percentage
  - Sampling percentage – 10
<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
</table>
| Eligible sample population          | (# of docs returned by the saved search - # of previously manually-coded docs) | Docs eligible for overturn # of 48,000:  
(50,000 - 2,000) = 48,000 |
| Sample set size                     | Sample set size (as determined by Sampling methodology):                | Sample set size: Statistical sampling sample set size of 5928:  
((1.64^2 * 0.25/1.0^2)/(1 + ((1.644853626951^2 * 0.25/1.0^2)-1)/47,950))) = 5,928 |
| Statistical Sampling:               | S = ceiling((V / ( 1 + ( (V – 1) / E ))) )                             | Percentage sampling sample set size of 4,800:  
(48,000 x 10)/100 = 4,800 |
| | Ceiling means to round up the calculated value to the next whole number |                                                                         |
| | V = see formula below               |                                                                         |
| | E = Docs eligible for overturn      |                                                                         |
| | V = (z^2 * 0.25) / c^2              |                                                                         |
| | z = statistics constant (see Calculated margin of error)                 |                                                                         |
| | 0.25 is used because RAR assumes the worst-case response distribution of 50%/50%, which gives the largest necessary sample size |                                                                         |
| | c = Desired margin of error percentage for richness:                     |                                                                         |
| | Percentage:                       |                                                                         |
| | (Eligible Sample Population x Sampling percentage field value)/100       |                                                                         |
| | Fixed sample size:                |                                                                         |
| | Value entered by user             |                                                                         |
| Overturned docs %                  | (Overturned count / Number of reviewed categorized documents)*100        | Overturned docs % of 100%:  
(150/1,500) x 100 |
| Calculated margin of error for richness | (sqrt (((pop – rc) * (z^2*0.25)) / (rc (pop-1))))*100               | Calculated margin of error value of 2.09:  
(√(((50,000-1,500) x (1.644853626951^2 x 0.25)/(1,500 (50,000-1)))) x 100 |
<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overturn range %</td>
<td>&lt;Lower overturn range % = Max(0, Overturned docs % - Calculated margin of error for richness)&gt; to &lt;Upper overturn range % = Min (100, Overturned docs % + Calculated margin of error for richness)&gt;</td>
<td>Estimated overturn range % range 8%-12%: 10.00 – 2.09 = 8% to 10.00 + 2.09 = 12%</td>
</tr>
<tr>
<td>Overturn range #</td>
<td>&lt;Lower overturn range # = Lower overturn range % x docs in round&gt; to &lt;Upper overturn range # = Upper overturn range % x docs in round&gt;</td>
<td>Estimated overturn range # 4,000-6,000: 8% x 50,000 = 4,000 to 12% x 50,000 = 6,000</td>
</tr>
</tbody>
</table>

### 5.3.4 Project Summary report

The **Project Summary** report provides a consolidated set of quality metrics from the entire project so that you can see the state of the project based on the last round completed.

**Note:** This report contains information from the last completed round in the project. This means that if a round in progress, it isn't reflected in this report.
Assisted Review: Project Summary Report

Project Name: Responsive Review
Project Status: Round complete. You can start your next round.
Last Completed Round: RESP 006
Project Creation Date: 6/20/2013
RAR Version: 9.3.411.10
Starting Project Universe: 16,642
Ending Project Universe: 16,642

Manually Coded Documents
The table below lists the number of documents that were manually coded by a human and the number of documents which were used as seed documents to train the system.

<table>
<thead>
<tr>
<th></th>
<th>Responsive (#)</th>
<th>Not Responsive (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coded</td>
<td>Seeds</td>
</tr>
<tr>
<td></td>
<td>2,564</td>
<td>2,188</td>
</tr>
</tbody>
</table>

Project Categorization Results
The table below lists the number and percent of documents from the ending project universe that were categorized by the system into each category along with those that were not able to be categorized.

<table>
<thead>
<tr>
<th>Categorized Responsive</th>
<th>Categorized Not Responsive</th>
<th>Uncategorized</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>7,192</td>
<td>43.22%</td>
<td></td>
</tr>
<tr>
<td>6,943</td>
<td>41.72%</td>
<td></td>
</tr>
<tr>
<td>2,507</td>
<td>15.06%</td>
<td></td>
</tr>
</tbody>
</table>

Project Overturn Results
The table below lists the percent of documents where the manually coded value disagrees with the system categorized value. The overturn range shows the estimate of overturns that would occur if every document in the latest round was manually reviewed.

<table>
<thead>
<tr>
<th>Responsive overturned docs (%)</th>
<th>Not Responsive overturned docs (%)</th>
<th>Overturned docs (%)</th>
<th>Estimated overturn range (%)</th>
<th>Confidence level</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.85%</td>
<td>3.75%</td>
<td>5.31%</td>
<td>2.76% - 7.86%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Control Set Statistics
The table below contains statistics that measure how well the control set documents were categorized in the last round. Precision tells you the percent of documents identified as responsive which are actually responsive. Recall tells you the percent of all the responsive documents that have been identified. F1 is a measure of accuracy and combines the precision and recall results.

<table>
<thead>
<tr>
<th>Richness</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Margin of error</td>
<td>Value</td>
<td>Margin of error</td>
</tr>
<tr>
<td>42.70%</td>
<td>±4.99%</td>
<td>74.17%</td>
<td>±6.55%</td>
</tr>
</tbody>
</table>

Coded Responsive (#) Coded Not Responsive (#)

<table>
<thead>
<tr>
<th>Categorized Responsive</th>
<th>Categorized Not Responsive</th>
<th>Uncategorized</th>
<th>Categorized Not Responsive</th>
<th>Categorized Responsive</th>
<th>Uncategorized</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>20</td>
<td>5</td>
<td>97</td>
<td>31</td>
<td>25</td>
</tr>
</tbody>
</table>
This report includes:

- **Manually Coded Documents** - the number of documents that were manually coded by a human along with the number of documents and document excerpts that were used as seeds to train the system.

- **Project Categorization Results** - the number and percent of documents from the ending project universe that were categorized by the system into each category along with those that were not eligible for categorization.

- **Project Overturn Results** - the percent of documents where the manually-coded value disagrees with the system-categorized value. The overturn range shows the estimate of overturns that would occur if every document in the last round was manually reviewed.

- **Control Set Statistics** - the statistics that measure how well the control set documents were categorized in the last round. This only appears if you’ve completed a control set round.

### 5.4 Control set reports

The Control Set Reports section on the Sample-Based Learning project console lists reports related to your control set documents.

**Note:** To access these reporting options, reviewers must have finished a control round during the project.

### 5.4.1 Understanding richness, precision, recall, and F1

The **Control Set Statistics** report shows the round-by-round statistics gathered for the control set of the project. Documents coded during a control set round are not included as examples but are instead used to determine the stability of a project by calculating richness, precision, recall, and F1.

- **Richness** - the percentage of documents in the control set reviewers coded as responsive.

- **Precision** - tells you the percent of documents categorized as responsive that reviewers coded as responsive (how accurate the results are). Specifically, it is the ratio of documents both categorized and coded as responsive to all of the documents that the system categorized as responsive in the control set.

- **Recall** - tells you the percent of documents reviewers coded as responsive that the system categorized as responsive (how complete the results are). Specifically, it is the ratio of documents both categorized and coded as responsive to all of the documents that reviewers coded as responsive in the control set.

- **F1** - tells you the measure of accuracy and combines the precision and recall results. Specifically, F1 is the harmonic mean or a weighted average of precision and recall.

Both precision and recall are based on an understanding and measure of relevance. This correlates to the positive chosen value in your project setup.
Sample-Based Learning arrives at precision and recall values via its determination of true and false positives, as well as false negatives:

- **True Positive** = coded responsive & categorized by the system as responsive
- **False Positive** = coded not responsive & categorized by the system as responsive
- **False Negative** = coded responsive & categorized by the system as not responsive or uncategorized

The following information illustrates how important measurements on the Control Set Statistics report are determined:

<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness</td>
<td>((N_r / N_{CS}) \times 100)</td>
</tr>
<tr>
<td></td>
<td>(N_{CS} = ) the control set size</td>
</tr>
<tr>
<td></td>
<td>(N_r = ) the number of documents in the control set that reviewers coded as responsive</td>
</tr>
<tr>
<td>Precision</td>
<td>Precision = True positive / True positive + False positive</td>
</tr>
<tr>
<td>Recall</td>
<td>Recall = True positive / True positive + False negative</td>
</tr>
<tr>
<td>F1</td>
<td>(2 \times ((\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})))</td>
</tr>
<tr>
<td>Name</td>
<td>Formula</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Calculated margin of error for richness</td>
<td>( z \times (\sqrt{\frac{(\text{pop} - \text{rc}) \times 0.25}{\text{pop} \times \text{rc} - 1})}) \times 100 )</td>
</tr>
<tr>
<td></td>
<td>- pop = docs in round</td>
</tr>
<tr>
<td></td>
<td>- rc = the number of reviewed categorized documents</td>
</tr>
<tr>
<td></td>
<td>- z = a statistical constant:</td>
</tr>
<tr>
<td></td>
<td>- 1.64 if confidence level = 90%</td>
</tr>
<tr>
<td></td>
<td>- 1.96 if confidence level = 95%</td>
</tr>
<tr>
<td></td>
<td>- 2.57 if confidence level = 99%</td>
</tr>
<tr>
<td>Calculated margin of error for precision</td>
<td>( z \times (\sqrt{\frac{(\text{pop} - N_{cs}) \times \text{precision} \times (1 - \text{precision} )}{\text{pop} \times (N_{cr} - 1})}) )</td>
</tr>
<tr>
<td></td>
<td>- pop = docs in round</td>
</tr>
<tr>
<td></td>
<td>- ( N_{cs} ) = the control set size</td>
</tr>
<tr>
<td></td>
<td>- ( N_{cr} ) = the number of documents in the control set that the system categorized as responsive</td>
</tr>
<tr>
<td></td>
<td>- precision = the calculated precision based on the control set documents</td>
</tr>
<tr>
<td></td>
<td>- z = a statistical constant:</td>
</tr>
<tr>
<td></td>
<td>- 1.64 if confidence level = 90%</td>
</tr>
<tr>
<td></td>
<td>- 1.96 if confidence level = 95%</td>
</tr>
<tr>
<td></td>
<td>- 2.57 if confidence level = 99%</td>
</tr>
<tr>
<td>Calculated margin of error for recall</td>
<td>( z \times (\sqrt{\frac{(\text{pop} - N_{cs}) \times \text{recall} \times (1 - \text{recall})}{\text{pop} \times (N_{r} - 1})}) )</td>
</tr>
<tr>
<td></td>
<td>- pop = docs in round</td>
</tr>
<tr>
<td></td>
<td>- ( N_{cs} ) = the control set size</td>
</tr>
<tr>
<td></td>
<td>- ( N_{r} ) = number of responsive documents in the control set</td>
</tr>
<tr>
<td></td>
<td>- recall = the calculated recall based on the control set documents.</td>
</tr>
<tr>
<td></td>
<td>- z = a statistical constant:</td>
</tr>
<tr>
<td></td>
<td>- 1.64 if confidence level = 90%</td>
</tr>
<tr>
<td></td>
<td>- 1.96 if confidence level = 95%</td>
</tr>
<tr>
<td></td>
<td>- 2.57 if confidence level = 99%</td>
</tr>
</tbody>
</table>

### 5.4.2 Control Set Statistics report

To access this report click **View Control Set Statistics** under **Control Set Reports** on the console.

The **Control Set Statistics** report shows the round-by-round statistics gathered for the control set of the project. Documents coded during a control set round are not included as examples but are instead used to determine the stability of a project by calculating richness, precision, recall, and F1. See **Understanding richness, precision, recall, and F1 on page 78** for more information on how these values are calculated in Relativity.
- **Richness** - the percentage of documents in the control set reviewers coded as responsive.

- **Precision** - tells you the percent of documents categorized as responsive that reviewers coded as responsive (how accurate the results are). Specifically, it is the ratio of documents both categorized and coded as responsive to all of the documents that the system categorized as responsive in the control set.

- **Recall** - tells you the percent of documents reviewers coded as responsive that the system categorized as responsive (how complete the results are). Specifically, it is the ratio of documents both categorized and coded as responsive to all of the documents that reviewers coded as responsive in the control set.

- **F1** - tells you the measure of accuracy and combines the precision and recall results. Specifically, F1 is the harmonic mean or a weighted average of precision and recall.

**Note:** If the number of documents in your project changes, the control set will no longer be representative of the entire project, and you will be notified of this occurrence on the Control Set Statistics report. In this case, you may want to create a new control set to represent all documents in the project.

The Control Set Statistics graph provides the following data:
- **Percent** - the graph’s vertical axis; this tracks the statistics of precision, recall, and F1 for the control set

- **Round** - the graph’s horizontal axis; this contains the control set rounds for which precision, recall, and F1 were calculated

The **Control Set Information** table includes information about the control set and how it was coded by reviewers.
### Control Set Information

<table>
<thead>
<tr>
<th>Richness</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>Margin of error</td>
<td>Estimate</td>
<td>Margin of error</td>
<td>Estimate</td>
</tr>
<tr>
<td>42.70%</td>
<td>±4.99%</td>
<td>74.17%</td>
<td>±9.74%</td>
<td>78.07%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control set round name</th>
<th>Reviewed docs in control set (#)</th>
<th>Coded Responsive</th>
<th>Coded Not Responsive</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESP 001</td>
<td>267</td>
<td>114</td>
<td>42.70%</td>
</tr>
</tbody>
</table>

This table provides the following data:

- **Richness** - identifies the estimated percentage of responsive documents for the entire control set round population and the margin of error percentage for that calculation.
- **Precision** - identifies the percentage of documents that Relativity correctly categorized as responsive in the control set round and the margin of error percentage for that calculation.
- **Recall** - identifies the percentage of documents that were categorized as responsive (out of all the responsive documents in the control set round) and the margin of error for that calculation.
- **F1** - the weighted average of precision and recall calculated in the round.
- **Confidence level** - the percentage for confidence level selected when the control set round was started.
- **Control set round name** - identifies which round is the control set.
- **Reviewed docs in control set (#)** - the number of documents in the control set round that were reviewed.
- **Coded responsive** - the number and percentage of documents coded as responsive in the control set round.
- **Coded Not Responsive** - the number and percentage of documents coded as not responsive in the control set round.

The **Control Set Statistics by Round** table includes precision, recall, and F1, as well as the distribution of how the control set documents were categorized for each round.

This table provides the following data:
## Control Set Statistics by Round

<table>
<thead>
<tr>
<th>Round name</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1 (%)</th>
<th>Categorized Responsive (%)</th>
<th>Categorized Not Responsive (%)</th>
<th>Uncategorized (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESP 001</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>100.00%</td>
</tr>
<tr>
<td>RESP 002</td>
<td>53.57%</td>
<td>13.16%</td>
<td>21.13%</td>
<td>10.49%</td>
<td>25.09%</td>
<td>64.42%</td>
</tr>
<tr>
<td>RESP 003</td>
<td>77.36%</td>
<td>71.93%</td>
<td>74.55%</td>
<td>39.70%</td>
<td>31.84%</td>
<td>28.48%</td>
</tr>
<tr>
<td>RESP 004</td>
<td>78.90%</td>
<td>75.44%</td>
<td>77.13%</td>
<td>40.82%</td>
<td>39.70%</td>
<td>19.48%</td>
</tr>
<tr>
<td>RESP 005</td>
<td>75.42%</td>
<td>78.07%</td>
<td>76.72%</td>
<td>44.19%</td>
<td>42.32%</td>
<td>13.48%</td>
</tr>
<tr>
<td>RESP 006</td>
<td>74.17%</td>
<td>78.07%</td>
<td>76.07%</td>
<td>44.94%</td>
<td>43.82%</td>
<td>11.24%</td>
</tr>
</tbody>
</table>

- **Round name** - the name of the round in which precision, recall, and F1 were calculated for the control set.
- **Precision (%)** - the percentage of documents identified as responsive which are actually responsive in the round. See the [Sample-Based Learning measurement descriptions](#) for additional details.
- **Recall (%)** - the percentage of responsive documents that were identified in the round. See the [Sample-Based Learning measurement descriptions](#) for additional details.
- **F1 (%)** - the weighted average of precision and recall calculated in the round. See the [Sample-Based Learning measurement descriptions](#) for additional details.
- **Categorized Responsive** - the percentage of documents in the round that were categorized as responsive.
- **Categorized Not Responsive** - the percentage of documents in the round that were categorized as not responsive.
- **Uncategorized** - the percentage of documents in the round that were not categorized.

The **Control Set Details by Round** table shows the numbers used in the precision, recall, and F1 calculations for each round.

This table provides the following data:

## Control Set Details by Round

<table>
<thead>
<tr>
<th>Round name</th>
<th>Coded Responsive (#)</th>
<th>Coded Not Responsive (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Categorized Responsive</td>
<td>Categorized Not Responsive</td>
</tr>
<tr>
<td>RESP 001</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>RESP 002</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>RESP 003</td>
<td>82</td>
<td>16</td>
</tr>
<tr>
<td>RESP 004</td>
<td>86</td>
<td>19</td>
</tr>
<tr>
<td>RESP 005</td>
<td>89</td>
<td>18</td>
</tr>
<tr>
<td>RESP 006</td>
<td>89</td>
<td>20</td>
</tr>
</tbody>
</table>
- **Round name** - the name of the round in which the control set documents were categorized.
- **Coded Responsive (#)** - provides the number of documents that were coded responsive which were also categorized responsive, categorized not responsive, or uncategorized.
- **Coded Not Responsive (#)** - provides the number of documents that were coded not responsive which were also categorized not responsive, categorized responsive, or uncategorized.

The **Summary of Rounds** table gives a breakdown of all settings made per round and provides the following data:

- **Round name** - the name of the round.
- **Round Type** - the type of round.
- **Reviewed docs in round (#)** - the number of documents reviewed in the round.
- **Reviewed docs in project (#)** - the total number of documents that have been reviewed in the project so far; this is a sum of all the values for the Reviewed docs in round (#).
- **Docs in round (#)** - the number of documents returned in the saved search for that round.
- **Round description** - the description entered when the round was started.

The **Control Set Category Volatility** graph and table display the number and percentage of documents that have changed designation category or become categorized per round. If you do not categorize when you finish a round, there will be no volatility for that round. This is useful because you can determine the stability of your categorization results by seeing how many control set documents are changing categories from round to round.

This graph provides the following data:
- **Percent Volatility** - the graph's vertical axis; this contains the percentage of documents that have changed category from one round to the next and become categorized per round.

- **Round** - the graph’s horizontal axis; this contains the rounds in which the categorization or category change occurred.

The **Control Set Category Volatility** table provides the following data:
- **Round name** - the name of the round for which control set volatility was calculated.
- **Not Responsive -> Responsive #** and % - the number and percentage of documents that changed from the not responsive category to the responsive category in that round.
- **Responsive -> Not Responsive #** and % - the number and percentage of documents that changed from the responsive category to the not responsive category in that round.
- **Uncategorized change #** and % - the number and percentage of documents that remained uncategorized in that round.

### 5.5 Issue reports

The Issue Reports section on the Assisted Review project console lists a variety of options for gathering insights into how well issues were categorized for your Assisted Review project.

**Note:** To access the issue reports, reviewers must have categorized for issues during the project.

#### 5.5.1 Issue Summary report

To access this report click **View Issue Summary** under **Issue Reports** on the console.

**Note:** To access the issue summary report, reviewers must have categorized for issues during the project.

The **Issue Summary** report provides a detailed breakdown of categorization results of issue coding after each round. The graph at the top of the report shows the most recent results.

The **Issue Categorization Results: <Round Name>** graph provides the following data:
- **Issue** - the graph's vertical axis; this contains each issue value and a bar representing the percentage of documents in the project that were categorized with that value.

- **% of docs in project** - the graph's horizontal axis; this contains percentage that correspond with the issue value bars in the vertical axis.

The **Issue Categorization Results: All Rounds** table provides the following data:
<table>
<thead>
<tr>
<th>Issue category</th>
<th>Round name</th>
<th>Manually coded Docs in round (#)</th>
<th>Categorized Docs in project (#)</th>
<th>Categorized Docs in project (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting</td>
<td>S85 001</td>
<td>3</td>
<td>3</td>
<td>0.03%</td>
</tr>
<tr>
<td></td>
<td>S85 002</td>
<td>0</td>
<td>3</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>S85 003</td>
<td>0</td>
<td>3</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>S85 004</td>
<td>0</td>
<td>3</td>
<td>0.09%</td>
</tr>
<tr>
<td>Energy Trading</td>
<td>S85 001</td>
<td>5</td>
<td>317</td>
<td>9.06%</td>
</tr>
<tr>
<td></td>
<td>S85 002</td>
<td>0</td>
<td>317</td>
<td>9.06%</td>
</tr>
<tr>
<td></td>
<td>S85 003</td>
<td>0</td>
<td>317</td>
<td>9.06%</td>
</tr>
<tr>
<td></td>
<td>S85 004</td>
<td>0</td>
<td>317</td>
<td>9.06%</td>
</tr>
<tr>
<td>Fraud</td>
<td>S85 001</td>
<td>1</td>
<td>2</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>S85 002</td>
<td>0</td>
<td>2</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>S85 003</td>
<td>0</td>
<td>2</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>S85 004</td>
<td>0</td>
<td>2</td>
<td>0.06%</td>
</tr>
<tr>
<td>Gas</td>
<td>S85 001</td>
<td>4</td>
<td>20</td>
<td>0.57%</td>
</tr>
<tr>
<td></td>
<td>S85 002</td>
<td>0</td>
<td>20</td>
<td>0.57%</td>
</tr>
<tr>
<td></td>
<td>S85 003</td>
<td>0</td>
<td>20</td>
<td>0.57%</td>
</tr>
<tr>
<td></td>
<td>S85 004</td>
<td>0</td>
<td>20</td>
<td>0.57%</td>
</tr>
<tr>
<td>HR Related</td>
<td>S85 001</td>
<td>2</td>
<td>3</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>S85 002</td>
<td>0</td>
<td>3</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>S85 003</td>
<td>0</td>
<td>3</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>S85 004</td>
<td>0</td>
<td>3</td>
<td>0.09%</td>
</tr>
</tbody>
</table>

- **Issue category** - the category created by the issue used to code documents.
- **Round** - the round in which the issue was applied
- **Coded docs in round (#)** - the number of documents coded with the issue value in that round
- **Categorized docs in project (#)** - the number of documents in the project that were categorized with the issue value in that round
- **Categorized docs in project (%)** - the percentage of documents in the project that were categorized with the issue value in that round

The data in the **Summary of Rounds** table provides the following data:
- **Round name** - the name of the round, as generated automatically when the round was started
- **Round type** - the type of round specified when the round was started: training, quality control, pre-coded seeds, or control set
- **Docs in project (#)** - the total number of documents in the project
- **Round description** - the description of the round, as provided by the project manager when starting the round

### 5.5.2 Designation-Issue Comparison report

The **Designation-Issue Comparison** report shows where designation and issues rank intersect.
Note: This report is only available after you have categorized both designation and issues in your project.

To access this report, click View Designation-Issue Comparison under Issue Reports on the console. Note the following about this report:

- The size of each plot on the chart is determined by the number of documents that fall into each intersection.
- The graph is updated each time you categorize.
- The y-axis contains designation rank for each category, usually responsive and not responsive, which are separated by the x-axis.
- Documents that have not been categorized by designation are displayed in orange on the x-axis.
- The x-axis contains issues rank; documents can fall into more than one issue category, and the issue rank is the weighted sum of the rank for each issue.
- Documents that are not categorized at all (i.e., not categorized for both Designation and Issue) do not appear in the graph.

Note: The Designation-Issue Comparison Report displays invalid or skewed results if users explicitly tag documents as "No Issue" or similar. Documents that do not have key issues should simply not be coded for this field.

5.5.2.1 Using the designation-issue comparison report

Each plot on the chart contains a range of documents that you can retrieve via a saved search to better analyze the data in this graph. You can then use these documents to set up subsequent rounds for manual review.

In most cases, plots that occur in the following areas of the graph merit review in a subsequent round of the project:
**Note:** Each bubble on the Designation-Issue Comparison Report graph represents an instance of the RAR Comparison RDO. The size of each bubble on the graph is determined by the value of a whole number field on the RAR Comparison RDO called Magnitude.

The following table breaks down the saved search conditions used to retrieve different types of documents from the Designation-Issue Comparison report:
<table>
<thead>
<tr>
<th>Document group</th>
<th>Field</th>
<th>Operator</th>
<th>Value</th>
</tr>
</thead>
</table>
| High issues range and uncategorized for designation (orange dots on the 0 line of the graph) | RAR Comparison        | these conditions | - Field: Designation-issue rank::Project prefix  
  - Operator: any of these  
  - Value: &lt;Project prefix&gt;  
- Field: Designation-Issue rank::Issues rank min  
  - Operator: is greater than  
  - Value: &lt;desired minimum rank&gt;  
- RAR Comparison::Designation Category  
  - Operator: is not set |
| High issues and high unresponsiveness (lower right corner of the graph)       | RAR Comparison        | these conditions | - Field: Designation-issue rank::Project prefix  
  - Operator: any of these  
  - Value: &lt;Project prefix&gt;  
- Field: Designation-Issue rank::Issues rank min  
  - Operator: is greater than  
  - Value: &lt;desired minimum rank&gt;  
- Field: RAR Comparison::Designation Category  
  - Operator: is  
  - Value: Non Responsive (enter manually) |
<table>
<thead>
<tr>
<th>Document group</th>
<th>Field</th>
<th>Operator</th>
<th>Value</th>
</tr>
</thead>
</table>
| High issues and high responsiveness (upper right corner of the graph) | RAR Comparison | these conditions | - Field: Designation-issue rank::Project prefix  
  - Operator: any of these  
  - Value: <Project prefix>  
- Field: Designation-Issue rank::Issues rank min  
  - Operator: is greater than  
  - Value: <desired minimum rank>  
- Field: RAR Comparison::Designation Category  
  - Operator: is  
  - Value: Responsive (enter manually)  
- Field: Designation-Issue rank::Designation rank min  
  - Operator: is greater than  
  - Value: <desired minimum rank> |
| Low issues and high responsiveness (upper left corner of the graph) | RAR Comparison | these conditions | - Field: Designation-issue rank::Project prefix  
  - Operator: any of these  
  - Value: <Project prefix>  
- Field: Designation-Issue rank::Issues rank min  
  - Operator: is less than  
  - Value: <desired minimum rank>  
- Field: Designation-Issue rank::Designation rank min  
  - Operator: is greater than  
  - Value: <desired minimum rank>  
- Field: RAR Comparison::Designation Category  
  - Operator: is |
Included in the report is the issue importance table, which lists the relative importance of each issue represented in the report. See Modifying issue importance on page 34 for more information on how to weight issue importance.

<table>
<thead>
<tr>
<th>Importance</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Importance</td>
<td>Gas</td>
</tr>
<tr>
<td>Medium Importance</td>
<td>Fraud</td>
</tr>
<tr>
<td>Low Importance</td>
<td>Energy</td>
</tr>
</tbody>
</table>

**Note:** Assisted Review uses the weights of each issue importance value to calculate the total issue rank of a document. For example, if a document is categorized with three issues, one of high importance (500%), one of medium (100%), and one of low (10%) and each with a rank of 90, the issue sum calculation is: 500% * 90 + 100% * 90 + 10% * 90 = 549.

### 5.6 RAR project document reporting views

The Assisted Review project console also provides a way to drill down to the list of documents that you want to see after investigating reports to analyze your control set documents, overturned documents, or saved categorization results on a more granular level.

#### 5.6.1 Viewing control set documents

The **Control Set Documents** list provides information on how all documents included in the control set were categorized for every round in the project.

To access this list click **View Control Set Documents** under **Control Set Reports** on the console.

The Control Set Documents view includes the following fields:

- **Assisted Review Project** - project in which the control set occurred.
- **Control Set** - the set created when the user finished the control set round.
- **Round** - the round in which the control set documents were categorized.
- **Document** - the file included in the control set.
- **Manually applied code** - the coding value manually applied to the control set document during review.
- **Control set reviewer** - the name of the reviewer who coded the document.
- **Categorized value** - the value with which the document was categorized.
- **Category rank** - the measure of the document’s conceptual similarity to the seed document.
- **Seed document** - the example document closest conceptually to the control set document; this is used as the example for selecting the category for the control set document.
- **Seed excerpt** - the excerpt of text from the seed document; this is only populated if the excerpted text was the closest conceptual match to the control set document.
- **Seed reviewer** - the reviewer who most recently changed the seed document’s designation field.
- **Previous categorized value** - the value with which the document was categorized in the previous round.

### 5.6.1.1 Using the control set documents report

If you notice trends in the control set graphs or tables that require your attention, you can use the fields on the control set documents report to drill into documents that need a closer look.

For example, a number of documents in the Control Set Details table were categorized differently than their coding values. You can isolate these documents by filtering on the Round, Manually applied code, and Categorized value fields. You can then open these documents to investigate what in the content might have led to the discrepancy between coded and categorized values.

### 5.6.2 Viewing overturned documents

Several options exist for viewing overturned documents for your RAR project:

- [Overturned documents dashboard below](#)
- [Viewing control set documents on the previous page](#)

#### 5.6.2.1 Overturned documents dashboard

Every field on the Overturned Documents tab is available for Group By and Pivot On in the pivot grid and chart. You can create powerful pivot charts to look for trends in overturns. For example, pivot on Seed document to view which documents cause the most overturns or pivot on Overturning Reviewer to see who created the most overturns. Relativity Version introduces a built-in customizable **Overturned Documents** dashboard that automatically filters the item list of overturned documents for the most recent assisted review round and built-in pivot widgets that give you a simple way to visualize why documents may have been overturned. In addition, you can add your own custom pivot charts to visualize overturned document data in the most relevant way for your project. The **Overturned Documents** dashboard provides detailed information for all documents where coding values disagreed with the values the system applied.

To access this tab for an Assisted Review project, click **View Overturned Documents** under **Designation Reports** on the Assisted Review console for the selected project. The Overturned
Documents tab displays overturned documents automatically filtered for the most recent round that was executed. Built-in pivot charts display useful information surrounding the round.

Note: You must have the correct workspace permissions to add, edit, or save a dashboard and to save pivot widgets to your dashboard.

The following properties are listed for the overturned documents in the selected project and round:

- **Round overturned** - the round in which the document was overturned
- **Overturned document** - the document that was overturned
- **Manually applied code** - the coding value the reviewer gave to the overturned document
- **Overturning reviewer** - the reviewer who applied a different value from the original categorization value
- **Overturned category rank** - the measure of the document’s conceptual similarity to the seed document

Note: Document ranking is based on the proximity of documents with the Analytics index-defined concept space. Within each category, documents are ranked based on the conceptual relevance of each document to its closest example document (the seed) in that space.
- **Seed document** - the example document closest conceptually to the overturned document; this is used as the example for selecting the category for the overturned document

- **Seed excerpt** - the excerpt of text from the seed document that caused the overturn; this is only populated if the excerpted text was the closest conceptual match to the overturned document

- **Overtumed category** - the previous designation category of the overturned document before it was overturned. For example, if the manually applied code was responsive, the overturned category is always not responsive and vice versa; this is also the category of the seed document or excerpt.

- **Seed reviewer** - the reviewer who most recently changed the seed document’s designation field. It is recommended that you create a project prior to reviewing any documents. If you do not do this, Assisted Review operates as normally but the seed reviewer value is unknown. The Seed Reviewer field is populated if the following conditions are met:
  - The seed document’s document-coding field was reviewed after a project was created that uses the relevant doc coding field.
  - Mass edit was not used to code the seed document.

The following pivot charts are included by default:

- **Overtums by seed document** - bar chart depicting the top 15 seed documents and the count of documents that were overturned because of them. This enables you to see which seed documents caused the most overturns for the most recent round.

- **Overturned category by rank** - stacked bar chart that groups overturned documents by category rank, showing the breakdown for each category. This graph allows you to visually identify and focus on the overturned documents that are most conceptually similar to their seed document (the overturns with the highest rank).

- **Overturns by overturning reviewer** - stacked bar chart depicting the count of overturned documents resulting from each overturning reviewer for each category. This enables you to see which overturning reviewer created the most overturned documents and for which categories.

- **Overturns by seed reviewer** - stacked bar chart depicting the count of overturned documents resulting from each seed reviewer for each category. This enables you to see which seed reviewer created the most overturned documents and for which categories.

You may also find the following pivot useful to add:

- **Seed reviewers by overturned category rank** – Bar chart that depicts the count of overturned documents for each seed reviewer, organized by overturned category rank. This enables you to see which seed reviewers are getting overturned where the seed document and overturned document are conceptually very similar to each other.

**Note:** For more information on how to create your own custom pivot charts for this dashboard and how to use Relativity dashboard functionality, see Dashboards in the Admin Guide.

### 5.6.2.2 Reviewing overturned and seed documents in the Overtum Analysis viewer pane

You can view overturned documents, seed documents, and seed excerpts in the Overtum Analysis pane in the Viewer. The document you have open in the viewer is the document that’s highlighted in the
Overtur Analysis pane. To view overturms and seed documents click in the pane menu.

In the image above, document Email 0005 was the original seed document and the document that would appear in the viewer, Email 0542, was the overturned document.

Note: The RAR icon is only available if a project exists in the workspace and you have access to the Overtur object.

The Analysis Overtur pane contains the following information:

- **Control number** - the identifying number for the document (whatever you named this field)
- **RAR overturn status** - the most relevant overturn information for each document in the pane
- **Designation rank** - the rank for each overturned document. Seed documents always have a rank of 100.

### 5.6.3 Viewing saved categorization results

The Saved Cat. Results report provides details on each set of saved categorization results.

To access this report, click View Saved Categorization Results under Designation Reports on the console.

This report provides the following data:

- **Assisted Review Project** - the name of the project
- **Saved cat. Results Set** - the name entered for the Categorization results set name field when finishing the round
- **RAR Saved Document** - the name of the document
- **Category** - the categorization value the document was tagged with for designation and/or issue
- **Category Rank** - the number ranking of how close the document is conceptually to the seed document, which is what gave the Saved Documents its assigned category
- **Document Coding Field** - the field entered for the Designation or RAR issue field when you created the project
- **Seed Document** - the document originally coded and used by Analytics as an example for the Saved Document

- **Saved Cat. Results::Description** - the description entered for the Categorization results set when finishing the round

- **System Created On** - the date and time at which the categorization results set was created
6 Errors and audit history

Relativity records all errors encountered at any point in the Sample-Based Learning project. You can access the details of these errors or retry them any time the View Errors and Retry Errors options are enabled on the console.

Using errors

Let's say you're managing an Sample-Based Learning project that encountered an error resulting from categorization failure in a later round. You need to access this error and find out why the categorization failed. You also need to get the agent ID associated with the error in case you need to troubleshoot it.

To do this, click the View Errors button on the console. To find the specific error you're looking for, you can filter the error list first by project and then, if you know what round in which it occurred, by the round name. You can also filter for the keyword "categorization."

You easily locate your error, referring to the Audit Details field, you learn that the categorization failure in that round occurred because the selected Analytics index didn't have queries enabled. This is an easy fix that involves simply navigating to that Analytics index and enabling its queries. There is no need for you to do anything with the error's agent ID.

Once you enable queries for the index, you can retry the errors for this project.

6.1 Viewing errors

If errors occur during any round of your project, the View Errors option is enabled in the console after the round is over.

To view errors:
1. Click **View Errors** on the console.

2. Refer to the error information on the project layout and audit list. Relevant errors have a value of **Assisted Review error occurred** under the Audit Action column.
3. Click the name of an error to go to its audit information. View its stack trace in the Audit details field.

4. If you don’t understand what the error is, and/or the error remains after you attempt to resolve it through the Retry Errors option on the console, contact support@relativity.com for assistance.

6.2 Retrying errors

To attempt to resolve errors that occurred during the round:

1. Click Retry Errors. Click OK on the confirmation message. Clicking Retry errors does the following:
   - Change the status of the project to Retrying errors.
   - Disables all console buttons except Refresh Page and Audit History.
2. Click **Refresh Page** to see the status of the project during error resolution.

3. If the resolution attempt fails, the View Errors and Retry Errors options are re-enabled. You can again view an error's stack trace in its Audit Details field and/or attempt to retry the error again.

   **Note:** If an error occurs on an Analytics Categorization Set, the Retry Errors button is disabled. Go to the Categorization Set to resolve the error.

### 6.2.1 Audit history

The Assisted Review Audit report provides a breakdown of all audit actions and details for the entire project.

Click **View Audit History** to bring up this list.
This report provides the following fields:

- **Name** - links to more information about the audit.
- **Audit Action** - the action performed in the project. The following are audited:
  - Assisted Review upgraded
  - Categorization set created
  - Round initiated
  - Sample set created
  - Sample set size
  - Document review started
  - End manual review by reviewer
  - Finish round invoked
  - Finished round invoked prior to all documents reviewed in sample set
  - Categorization started
  - Categorization finished
  - Round finished
  - Assisted Review error occurred
  - Agent recovery information

<table>
<thead>
<tr>
<th>Name</th>
<th>Audit Action</th>
<th>Audit Details</th>
<th>Action By</th>
<th>Assisted Review Round</th>
<th>System Created On</th>
</tr>
</thead>
<tbody>
<tr>
<td>AssistedReviewAudit</td>
<td>Assisted Review Error Occurred</td>
<td>An error occurred: -PR2-RAA Designation Cat. Set- Categorization failed because the selected index does not have queries enabled. Please enable queries on the index and then retry errors on the categorization set(s) named with the project prefix.</td>
<td>System</td>
<td>P02 001</td>
<td>10/16/2014 16:27 AM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Agent Recovery information</td>
<td>The recovery process is complete, and the action will retry. Actions taken in sequential order. The project status changed to Finished Round.</td>
<td>System</td>
<td>P02 001</td>
<td>10/16/2014 16:27 AM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Agent Recovery information</td>
<td>Recovery Process begun for project 1042542</td>
<td>System</td>
<td>P02 001</td>
<td>10/16/2014 16:27 AM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Assisted Review Error Occurred</td>
<td>An error occurred: -PR2-RAA Designation Cat. Set- Categorization failed because the selected index does not have queries enabled. Please enable queries on the index and then retry errors on the categorization set(s) named with the project prefix.</td>
<td>System</td>
<td>P02 001</td>
<td>10/15/2014 2:00 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Categorization Started</td>
<td>P02 001, C/21 2:52:252A</td>
<td>System</td>
<td>P02 001</td>
<td>10/15/2014 2:00 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Finish Round Invoked prior to all documents reviewed in sample set</td>
<td>P02 001</td>
<td>Smith, Jane</td>
<td>P02 001</td>
<td>10/15/2014 2:00 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Document Review Started</td>
<td>P02 001</td>
<td>Smith, Jane</td>
<td>P02 001</td>
<td>10/15/2014 1:49 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Batches Created</td>
<td>RAR P02 001</td>
<td>Smith, Jane</td>
<td>P02 001</td>
<td>10/15/2014 1:45 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Sample Set Size</td>
<td>82</td>
<td>System</td>
<td>P02 001</td>
<td>10/15/2014 1:45 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Sample Set Created</td>
<td>P02 001</td>
<td>System</td>
<td>P02 001</td>
<td>10/15/2014 1:45 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Round Initiated</td>
<td>P02 001</td>
<td>Smith, Jane</td>
<td>P02 001</td>
<td>10/15/2014 1:45 PM CDT</td>
</tr>
<tr>
<td>AssistedReviewAudit</td>
<td>Categorization Set Created</td>
<td>P02-RAA Designation Cat. Set</td>
<td>System</td>
<td>P02 001</td>
<td>10/15/2014 1:44 PM CDT</td>
</tr>
</tbody>
</table>
- **Audit Details** - the corresponding detail to the audit action. For example, if the action is Round initiated, the audit detail will be the name of that round.

- **Action By** - the name of the user who performed the action. If the action was automatically performed by the system, then the value for this field will be system.

- **Assisted Review Round** - the round in which the audit action occurred.

- **System Created On** - the date and time at which the audit action was performed.
Proprietary Rights

This documentation ("Documentation") and the software to which it relates ("Software") belongs to Relativity ODA LLC and/or Relativity’s third party software vendors. Relativity grants written license agreements which contain restrictions. All parties accessing the Documentation or Software must: respect proprietary rights of Relativity and third parties; comply with your organization’s license agreement, including but not limited to license restrictions on use, copying, modifications, reverse engineering, and derivative products; and refrain from any misuse or misappropriation of this Documentation or Software in whole or in part. The Software and Documentation is protected by the Copyright Act of 1976, as amended, and the Software code is protected by the Illinois Trade Secrets Act. Violations can involve substantial civil liabilities, exemplary damages, and criminal penalties, including fines and possible imprisonment.

©2019. Relativity ODA LLC. All rights reserved. Relativity® is a registered trademark of Relativity ODA LLC.