

Guide to making assessments on the single trace platform

INTRODUCTION	1
The repliCATS project	2
The IDEA protocol	2
Guidelines for Round 1	
Entering your Round 1 estimates	3
Answering the questions	4
Question 1 (comprehension)	4
Answering the question	4
Question 2 (plausibility)	5
Answering the question	5
Question 3 (replicability)	6
Answering the question	6
Understanding the three-step format	6
Things to consider	7
Question 4 (other aspects)	8
Purpose	8
Clarification	8
Question 5 (involvement)	8
Guidelines for Round 2	9
Entering your Round 2 estimates	9
Guidelines for a good Discussion phase	9
Ground rules	9
Tips for a good Discussion phase	10
Help and FAQs	10
I feel unqualified to answer these claims	10
I'd like to access some training materials and practice questions	11
I don't understand what is meant by 'replication' and other terms	11
I am involved in a replication study for a claim that I have been assigned	11
Will my data be made publicly available? Can I withdraw it?	11
Code of conduct	11
Who can I contact about...	11
Who can I contact if I have concerns about the project itself?	11

Introduction

This document contains a quick reminder of the aims and approach of the repliCATS project, followed by a guide to answering the claims in Round 1 and Round 2 of the IDEA protocol. There is also a list of FAQs.

The repliCATS project

The University of Melbourne repliCATS team elicits expert judgements about the credibility of research claims in the Social and Behavioural Sciences through an online platform using the IDEA protocol. Judgements are aggregated into measures of reliability and the reasoning used is analysed. IDEA (“Investigate”, “Discuss”, “Estimate” and “Aggregate”) has been found to improve judgements under uncertainty. More information about the repliCATS project is contained in the [Plain Language Statement](#) and on the [repliCATS website](#).

The IDEA protocol

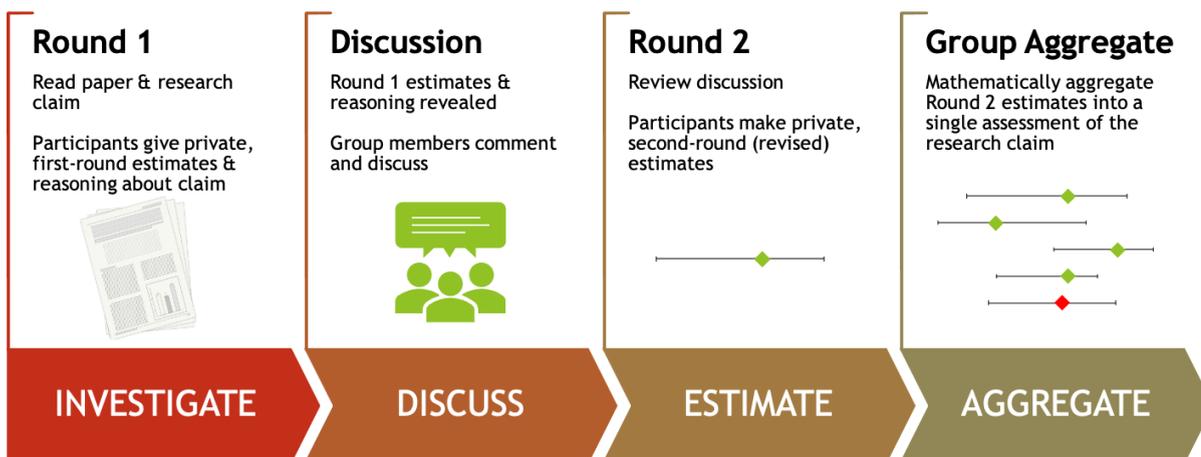
This protocol, developed at the University of Melbourne, has been found to improve judgements under uncertainty. IDEA stands for “Investigate”, “Discuss”, “Estimate” and “Aggregate”, the four steps in the process of this elicitation.

As used in the repliCATS project, the IDEA protocol will involve participants:

- 1 Independently *Investigating* the claim, providing their personal judgement on the credibility of the claim, and commenting on their thinking.
- 2 Seeing the judgements of the rest of their team, the aggregated judgement and all of the comments that have been made and having a *Discussion* with the group. This phase can resolve uncertainties and investigate evidence and thinking.
- 3 Providing a revised *Estimate* and describing how their thinking has changed.

The repliCATS team will use an *Aggregate* of the group judgements as the final assessment of the replicability of the research claim.

More information on the IDEA protocol can be found [here](#).



Guidelines for Round 1

Entering your Round 1 estimates

- a) **Login** to the repliCATS platform at <https://replicats-singletrace.eresearch.unimelb.edu.au>. If you have difficulty, please contact replicats-contact@unimelb.edu.au. The platform is tested for Google Chrome on laptops. It may not work in incognito mode, on other browsers, or on mobile devices.
- b) **Select a claim.** Upon login you see the list of claims that have been assigned to you. If you are in a facilitated session, your facilitator will advise you about claims to select. Otherwise, select any claim by clicking on it. You'll be taken to the claim assessment page with information about the claim and the elicitation questions (outlined below).
- c) **Investigate the claim.** We don't expect you to take more than about 5-15 minutes investigating the claim. Some information about the claim has been extracted, including the abstract and basic statistics. You also have a link to the original paper. The claims have been selected by a team independent to repliCATS.
We encourage you to use whatever information you would like to inform your judgements. You can speak to others and search for additional sources of evidence. However, if you know someone is assessing the same claim as you, please **do not** discuss your judgements with them – you can do this in the *Discussion* phase.
- d) **Step 3: Answer the questions.** Details of these questions are provided below. We don't expect you to spend longer than 5-15 minutes per claim answering the questions. You will have time to revise your estimates in Round 2.
We encourage you to use the comments boxes provided to write notes to help you remember crucial aspects that informed your responses or matters that you'd like to clarify in the Discussion phase. We are also interested in understanding how you thought about the claim and what characteristics were important for your judgement. Note that all comments will be attributed to an avatar. **Please do not use your own name or the name of any other participants.** Judgements will eventually be made public and we must be able to keep them anonymous.
- e) **Save and submit your estimates.** 'Saving' your estimates will enable you to return to the estimates and update them prior to submission. Pressing 'submit' finalises your estimates for Round 1 - but you'll have an opportunity to revise these in Round 2 following the Discussion phase. Check to see that the claim is listed as 'Round 2' on the home page. If not, you will need to click back into the claim and re-submit.
- f) **Enter your Round 2 judgements** if you are working remotely online, i.e. not in a facilitated session, we would like you to enter a judgement in Round 2 straightaway, even if there are no other Round 2 judgements to consider. This will enable us to process the claims more quickly. You can always update your Round 2 judgements at any time before the claim is finalized – you will be notified before this happens. More details are available in our Guide to Round 2. If you are working in a facilitated session, your facilitator will tell you when to move to Round 2.

That's Round 1 completed. You can also see the basic steps in completing Round 1 on the platform in a video that can be found [here](#).

Answering the questions

For each claim there are four main questions that we ask you to answer (plus a fifth, that has a standard response). Some questions may appear to be asking for similar information, but they do have different purposes. In this section we outline why we are asking each question, and the reason for the formats chosen.

Question 1 (comprehension)

How well do you understand this claim?

Purpose: To understand if anything affects your ability to interpret the claim.

Clarification: We've done our best to make sure the claims are interpretable but we know they vary in clarity and comprehensibility. We've extracted what statistical data we can from the article to help inform your judgement. However, the process is automated and is not perfect. It's possible that the claim:

- is vague;
- is poorly written;
- relies on an unfamiliar procedure;
- contains too much jargon;
- has differences between the claim in the abstract and the claim which is tested; or
- is about a concept that you have difficulty conceptualising.

These factors can all contribute to your ability to be able to interpret the claim and may in turn lead to different interpretations by the group. There is a comments box below this question, where you can provide a summary of your interpretation of the question – this will be really useful for reminding you of what your interpretation was in the Discussion phase.

Answering the question

We're asking this on a scale. At the left end you can mark that you have no idea what the claim means, while the right end means it's perfectly clear to you.

Before answering try your best to interpret what you believe the claim might be asking. Some claims may be outside of your immediate knowledge or have some words you are unfamiliar with. This might cause you to immediately put 'I have no idea what the claim means'. However, with a little bit of effort you can usually deduce what is being asked either from the abstract or looking at the full-text publication. If after trying your best to understand the claim (i.e. 5 minutes of work), you still cannot work out what the claim is saying then you should definitely indicate this to us.

Comments box.

We have provided a comment box below the question so you can try to rephrase what you think the claim is asking. This will be really useful in the Discussion phase to prompt your memory about your initial interpretation of the question.

The placeholder text here (and in other questions) gives some suggestions about how you might answer, but they are only suggestions – you are always free to answer how you choose. The placeholder text for this question is:

You can restate the claim in clearer terms here. You could also highlight terms or concepts that you didn't understand. You have limited space to do this here, but don't worry, in question 4 below you can tell us any other thoughts you may have.

Question 2 (plausibility)

What's your initial reaction: is the underlying effect or relationship plausible?

Purpose: To capture your beliefs about whether the underlying effect or relationship corresponds to something real.

Clarification: Sometimes we hear a claim and we have a strong feeling that the claim being made does not seem very plausible either within the context of the experimental design, or more broadly (i.e. relating to a relationship that would generalise across contexts, or experimental designs).

These prior beliefs can be useful. We've included this question here to allow you to state your prior belief about if you think there is a real effect in this study, regardless of what you think about this particular experimental/study design.

Don't spend too much time on this question. In the next question, we want you to examine the claim and the validity of your prior beliefs more critically, as to how they relate to direct replication.

Answering the question

This is a binary question. Please select 'Yes' (if you believe the claim might relate to some underlying effect or relationship to something real) or 'No' (if you don't believe this).

The word 'plausible' means different things to different people. For some people almost everything is 'plausible', while other people have a stricter interpretation. Don't be too focused on the precise meaning of 'plausible' – you could also consider words like 'possible' or 'realistic' here. We just ask you to maintain a consistent standard between different claims and try to let us know if some claims are clearly *more* plausible (or implausible) than others.

If you didn't understand the claim being asked, it might be challenging to say whether you believe it's plausible. Please do your best to select an option but note this in the comments box in Question 4. Hopefully the claim will become clearer in the Discussion phase.

Comments Box: There is no comments box provided. However, if you have any strong beliefs or feelings about this claim, particularly why it may not be a real effect we'd like to hear them. These are likely to provide really important concepts that should be considered before trusting this claim more broadly. Please use the comments box in Question 4 to list some brief thoughts.

Question 3 (replicability)

What is the probability that close replications of this study would find a statistically significant effect in the same direction as the original claim (0-100%)?

Purpose

The question is asking about **replication**.

Clarification

- **A close replication** is a new experiment that follows the methods of the original study with a high degree of similarity, varying only aspects where there is a high degree of confidence that they are not relevant to the research claim. People often use the term direct replication – however no replication is perfectly direct, and we *cannot* describe precisely how any given claim will be replicated. Decisions about how to perform replications are made by a team of researchers that is independent from the replicATS project. Our best advice is to imagine what kinds of decisions you would face if you were asked to replicate this study, and then to consider the effects of making different choices for these decisions.
- **A successful direct replication** is one that finds a **statistically significant effect** (defined with an alpha of 0.05) that is in the same direction as the original study, using the same statistical technique as the original study. Assume that the direct replication has a sample size that is at least as large as the original study and has a goal of at least 95% power to detect the original study effect size.

Answering the question

In this question, we want you to try and think of reasons why the claim may or may not replicate. We understand that your thoughts about prior plausibility of the general claim is likely to influence your judgement regarding this question. However, we'd like you to try and think more critically of other reasons why this particular study may (or may not) replicate.

Sometimes the bold text from the abstract does not match the claim in the inferential test results. *In this case, your answers should relate to the inferential test results*, as the next stage of SCORE will focus on testing the replicability of the test results only.

Please only use whole integers for this question. Do not use decimal places.

Understanding the three-step format

The question asks you to provide three estimates of replicability: a lower bound, upper bound and best estimate. Here's how we want you to think of those questions:

- First, consider all the possible reasons **why a claim is unlikely to successfully replicate**. Use these to provide your estimate of the lowest probability of replication.
- Second, consider the possible reasons **why a claim is likely to successfully replicate**. Use these to provide an estimate of the highest probability of replication.
- Third, consider the balance of evidence. Provide your best estimate of the probability that a study will successfully replicate.

There is evidence that asking you to consider your lower and upper bounds before making your best estimate improves the accuracy of your best estimate. The difference between your upper and lower estimates is intended to reflect your uncertainty about whether that claim's findings would replicate. There's no 'correct' answer here, but we expect that your intervals for those claims you feel most uncertain about will be the widest.

Things to consider

There are many things you might consider when making your judgement. The IDEA protocol operates well when a diversity of approaches is combined. The following section lists some things you *may* wish to consider.

- **The statistical data, analyses and results for the claim**, including sample size, effect size and p value, if reported. These details are likely to be important for whether a claim replicates - see [this document](#) for more information.

There are many other important things you can investigate and consider.

- **The experimental design**. Will it be reliable in replication? Are there any signs of Questionable Research Practices e.g. unusual designs where more straightforward tests might have been run but failed? Note that this question is interested in the replicability of the claim even if the validity of the design is low.
- **Your prior plausibility** for the claim. Background probabilities are often a major factor. Is this area of research more or less well-understood?
- **Contextual information** about the original study or publication such as where and when the claim was published, and who undertook the original study. Do you have any private or personal knowledge e.g. experience with undertaking similar research, or existing knowledge about the quality of work from a particular source?

Some additional things to consider:

- Providing a lower estimate of 0 means you believe a study would never successfully replicate, not even by chance (i.e. you are certain it will not replicate). Providing an upper estimate of 100 means you believe a study would never fail to replicate, not even by chance (i.e. you are certain it will replicate).
- Providing an estimate of 50 means that you believe the weight of evidence is such that it is as likely as not that a study will successfully replicate. If you have low prior knowledge and/or large uncertainty, please use the width of your bounds to reflect this, and still provide your best estimate of the probability of replication.
- Answers **above 50** indicate that you believe it's more likely that the study would replicate than it would not replicate. Answers **below 50** indicate that you believe it's more likely that the study would not replicate than it would replicate. Intervals (the range between your lowest and highest estimate) which extend above and below 50 indicate that you believe there are reasons both for and against the study replicating.

Comments box: We ask you to provide a justification for your responses that can help your own thinking, and your team's discussion. Don't feel that you have to express these thoughts in polished prose. Dot points and partial notes are fine, as long as you can be understood. Any such notes will improve the Discussion phase, and the later analysis.

The placeholder text for this question is:

What factors influenced your judgement? What, if anything, made you think that it might successfully replicate? What, if anything, made you think that it might not successfully replicate?

Question 4 (other aspects)

“Considering the major factors that influenced your thinking in making these judgements, please describe any important aspects that you have not already covered above”.

Purpose

To obtain additional thoughts and comments.

Clarification

This question allows you to express any additional comments and information you have about the claim that you don't think fit well under previous comments sections. You can also describe how you went about reaching a judgement. We are interested in both your judgements and the way you went about making them. Your answer here gives an opportunity to inform our team and the discussion about the basis of your judgements. As with the previous comments box, partially formed thoughts or dot points are fine. You might even have contradictory opinions about the study – it's fine to express those as well.

The placeholder text for this question is:

You should answer this freely. You might describe the most relevant, important or interesting aspects of the discussion, and how these influenced your Round 2 judgements. This comment will not be displayed. If you have information or thoughts about this claim that you were not comfortable sharing with the group, this is an opportunity to include them.

Question 5 (involvement)

Were you involved in the writing, data collection, or analysis of the original study?

Purpose: We will interpret your estimates and reasoning differently if you are an author of a paper, or otherwise involved. We understand that you will probably always answer 'No' to this question.

Guidelines for Round 2

Entering your Round 2 estimates

Please enter some Round 2 estimates as soon as possible after completing Round 1 *even if you are the first to assess a question and there are no other responses to consider*. Claims start closing off once five participants have entered a Round 1 judgement – so if everyone holds off on Round 2 then we won't have a Round 2!

Once a claim starts closing, you will receive an email and will have 72 hours to update your estimates and reasoning. (These are the 'Pending to Close' claims on the platform.) Anyone with a Round 1 judgement can update at any time in this period, and as many times as they like. However, people who have not entered a Round 1 assessment for the claim cannot join.

Please do take the time to check the claim when you get the email – at this point everyone's assessments are in and the Discussion phase can really begin. Leaving questions and comments as soon as possible after you receive the email is extremely helpful, as is responding to them.

Please also consider whether you want to update your estimates or reasoning based on the Discussion phase. This is at the heart of the IDEA protocol. Remember, you have 72 hours – only – to update your estimates after you receive the notification that a claim is closing.

Guidelines for a good Discussion phase

The Discussion phase in the IDEA protocol is important. It provides an opportunity to resolve differences in interpretation, and to share and examine evidence.

When you are assessing claims in a facilitated workshop, where all participants are present at the same time, you can directly discuss the claim and ask questions of each other. Mostly, you will be assessing claims in a remote, online fashion and will need to use the comments features of the repliCATS platforms to interrogate other participants' reasoning and to ask questions of each other. Please do leave questions and comments for other participants on their reasoning, and consider and respond to the questions and comments on yours.

It is important that you do not use participant names on the platform, not even your own.

Reasoning comments will eventually be made public, and we need to keep these anonymous. If you want to refer to other participants, please use the anonymous handle they have been assigned e.g. Koala11.

Ground rules

Some ground rules for the Discussion phase, regardless of whether you are leaving comments on the online platform or discussing directly:

- Respect that the group is composed of a diversity of individuals.
- Consider all perspectives in the group. In synchronous discussion, allow an opportunity for everyone to speak.
- Don't assume everyone has read the same papers or has your skills – explain your reasoning in plain language.
- If someone questions your reasons, they are usually trying to increase their own understanding. Try to provide simple and clear justifications.
- Try to be open-minded about new ideas and evidence.

Tips for a good Discussion phase

The following list may be useful to consider when reviewing and commenting on judgements. You do not have to work through these systematically, but consider which may be relevant:

- What did people believe the claim being made was? Was the claim clear? Did everyone understand the information and terms in the same way?
- Consider the range of estimates in the group and ask questions about extreme values. What would cause someone to provide a high/low estimate for this question?
- Very wide intervals suggest unconfident responses. Are these based in uncertainties of interpretation or are these participants aware of contradictory evidence?
- Very narrow intervals suggest very confident responses. Do those participants have extra information?
- It's ok if you don't have good evidence for your beliefs – please feel free to state this.
- If you have changed your mind since your Round 1 estimates it's good to share this. Actively entertaining counterfactual evidence and beliefs improves your judgements.
- If you disagree with the group that is fine. Please state your true belief when completing Round 2 estimates. This represents the true uncertainty regarding the question, and it should be captured.
- Consider raising counterarguments, not to be a nuisance, but to make sure the group considers the full range of evidence.

Help and FAQs

We understand the task assigned to you is not easy, here are some tips to help you through.

I feel unqualified to answer these claims

There will be claims which you feel you are unqualified to answer. This is natural, the task is difficult (the purpose of this experiment is to understand whether it is even possible to systematically predict replicability). We ask you to please attempt all claims assigned. We do not expect you to have specific expertise in every claim you assess. We expect that both experts and non-experts contribute to good judgements. People who consider themselves 'outsiders' may notice details about the claim which more experienced members of the group overlook. However, if you genuinely cannot understand a claim, it will not be a satisfying experience to assess it. Thus, our standard is that you *feel comfortable* interpreting a claim, **not** that you *feel expert* in the subject. Remember:

- You can adjust your upper and lower bounds to express your uncertainty.
- You can draw on thoughts and opinions of colleagues (as long as they are not participants answering the same claims), or even additional resources.
- You can also justify your responses and any evidence or questions you have about the claim in the comments box.
- In Round 2 you'll be able to draw on the knowledge of the other participants and update your response in light of the Discussion phase.
- If you feel you really don't understand the claim, then please note this in Question 1, and provide your best interpretation of the claim's meaning. We can discuss these interpretations in the Discussion phase.

I'd like to access some training materials and practice questions

We have also developed some training which might help you to better assess the claims. This includes a downloadable training document that covers statistical concepts and some background information on questionable research practices and replication rates in previous studies. This training document can be downloaded [here](#). We will also send you an interactive quiz and example claims to practice on. Including both the document and quiz, the training takes about one hour to complete.

I don't understand what is meant by 'replication' and other terms

A list of terms can be found in our [glossary](#). If there is a term missing, please notify us.

I am involved in a replication study for a claim that I have been assigned

In the unlikely event that you are involved in the replication of one or more of the claims assigned, then please do not assess that claim, and please do not reveal which claim you are involved in replicating.

Will my data be made publicly available? Can I withdraw it?

At the end of this project, data will be made publicly accessible in an anonymized format. It is for this reason that you ***must not use the name of any participant including your own***. While you can end your participation at any time, we may not be able to remove data previously submitted to us.

Code of conduct

The repliCATS project has a strict [code of conduct](#). Please report any suspected breach.

Who can I contact about...

General repliCATS administration: Mel Ross repliCATS-contact@unimelb.edu.au

Time cards & salary: Mel Ross & Cassie Watts repliCATS-contact@unimelb.edu.au

Trouble logging in: Mel Ross repliCATS-contact@unimelb.edu.au

Other platform troubleshooting: Report through platform

Weekly group meetings: Your group leaders – Andy Head, Dan Hamilton or Rose O'Dea

General questions about assessing claims: Refer to participant materials

If still unresolved, send query to repliCATS-contact@unimelb.edu.au

Code of conduct breaches: Report in platform or to Fallon Mody fallon.mody@unimelb.edu.au

Who can I contact if I have concerns about the project itself?

This research project has been approved by the Human Research Ethics Committee of The University of Melbourne. If you have any concerns or complaints about the conduct of this research project, which you do not wish to discuss with the research team, you should contact the Manager, Human Research Ethics, Research Ethics and Integrity, University of Melbourne, VIC 3010. Tel: +61 3 8344 2073 or Email: humanethics-complaints@unimelb.edu.au. All complaints will be treated confidentially. In any correspondence please provide the name of the research team (provided above) and ethics ID number of the research project 1853445.