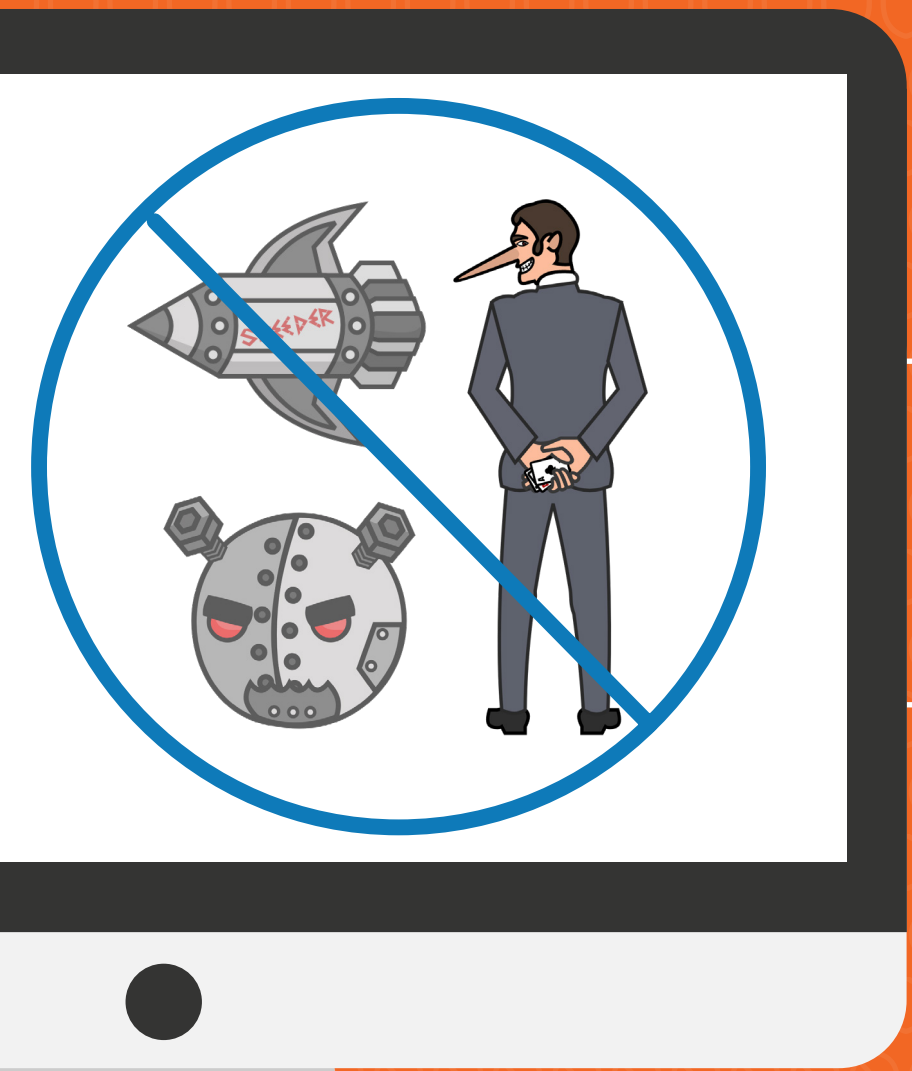
PREVENTING BAD DATA



How to Catch Survey Speeders, Cheaters & Bots

jibunu

Introduction

PAGE 4

What Are Speeders, Cheaters, and Bots?

PAGES 5-6

Why Are Speeders, Cheaters, and Bots Bad?

PAGES 7-8

Catching Speeders, Cheaters, and Bots

PAGES 9-11

Cleaning Speeders, Cheaters, and Bots, from Your Data

PAGES 12-16

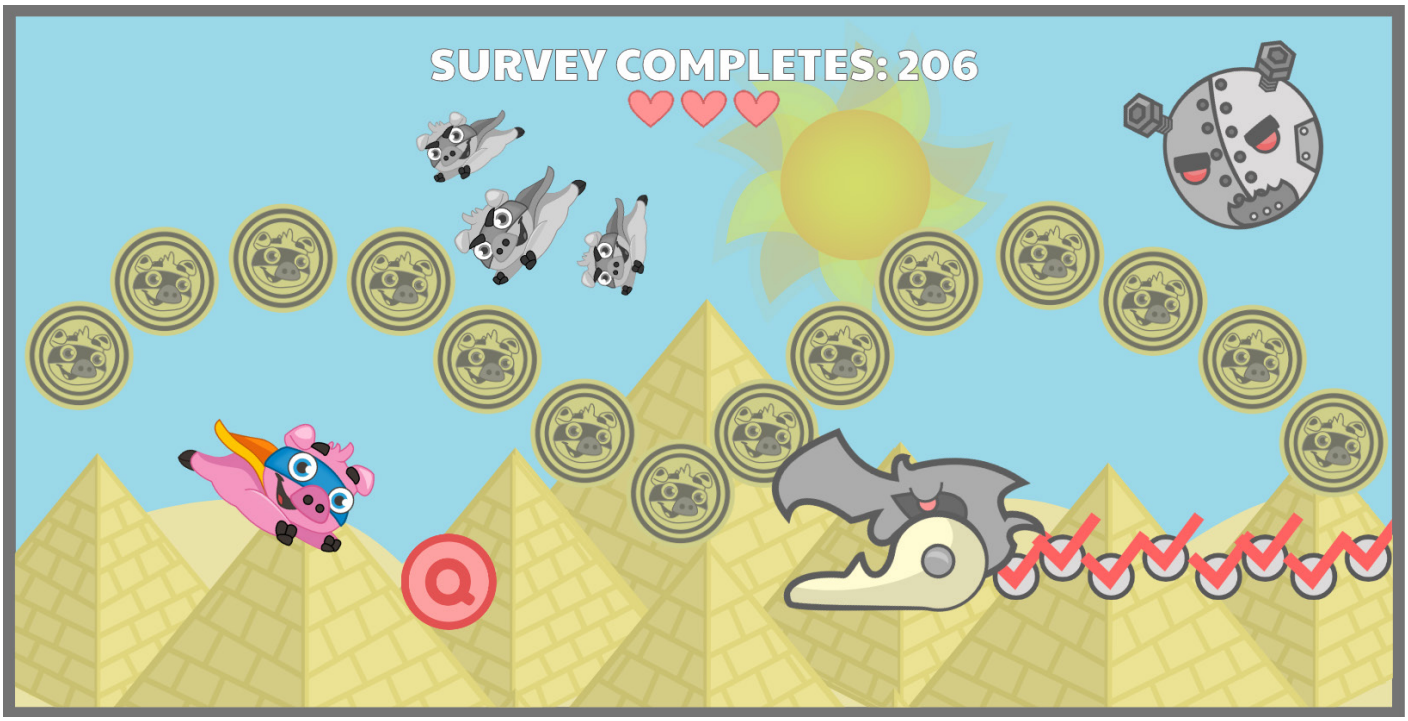
Preventing Speeders, Cheaters, and Bots

PAGES 17-20

Conclusion

PAGE 21

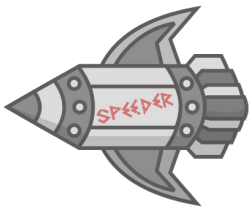
Save The World From Bad Data!



Hamlet's Survey Scramble

Play our market research related video game, Hamlet's Survey Scramble where you'll collect survey completes while avoiding the bad data mentioned in this eBook!

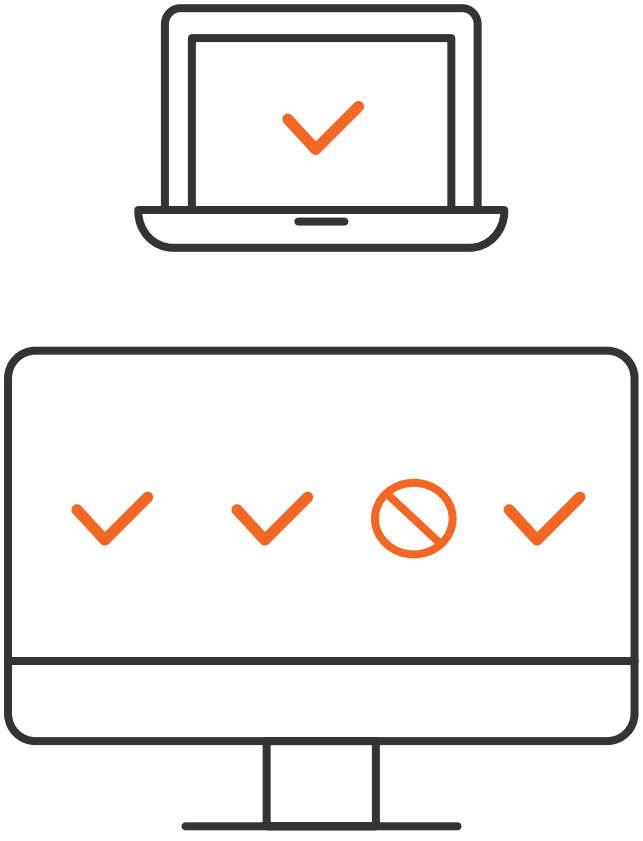
Click Here To Play

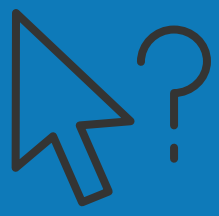


Introduction

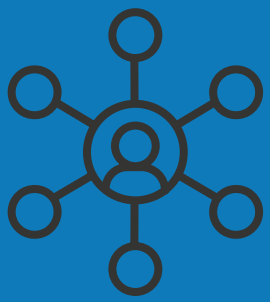
Most people taking surveys do so honestly, but there's always going to be some bad apples in the batch that you'll have to weed out. With online surveys having increased numbers of accepted respondents, bad data will make its way through the cracks. From Speeders to Cheaters, and everything in between, bad data can invade your survey and distort your data with inaccurate answers!

Thankfully, there are ways to counteract the scourge of survey villains. With a few key tricks, you can effectively push back against these problems and ensure that the integrity of your surveys remains intact.





What Are Speeders, Cheaters, and Bots?



SPEEDERS

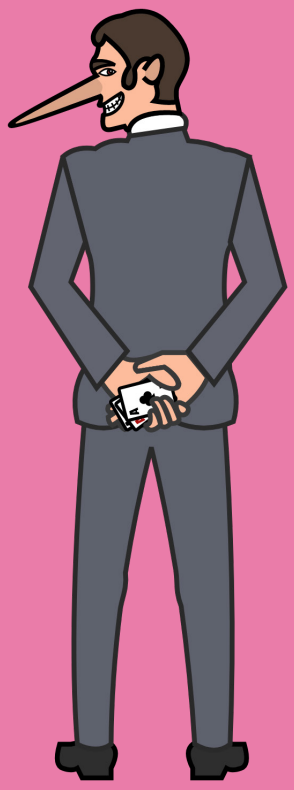
A speeder is, well, exactly what it sounds like; a respondent who rushes through your survey.

Seeing someone that finishes a survey fast doesn't guarantee that you have a speeder on your hands, and **finishing a survey quickly is not always a bad thing.** There is some research that shows that responding to questions quickly uses more of a person's System 1 processing and can be more in tune with how they really feel.

In his book *Thinking, Fast and Slow*, Daniel Kahneman, among other things, describes two systems of thinking. System 1 is fast, automatic, and emotional while System 2 is slow, effortful and calculating. So, depending on what your survey is about, you might want quick, System 1, responses or even encourage emotional responses by employing a time limit on certain questions. All that being said, there is a point where the respondent is going so fast that they are not thinking in System 1, but rather not thinking at all.

CHEATERS

Cheaters are respondents that give improper or incorrect answers in an effort to qualify for your survey, get through your survey, or save time when trying to get to the end of the survey. If there is an incentive offered, cheaters may also try to take the same survey over and over again hoping to cash in as many times as possible.



BOTS

Online bots have many uses. Chatbots allow websites to answer consumers frequently asked questions or lead users to proper content automatically. Testing bots can be used to test websites for broken links or other potential issues. Survey bots can even be used in a positive way to create a test data set and check a large number of paths within a survey automatically. It's when this bot technology is used to cheat the system and gain the rewards for completing surveys that we need to worry about them.

Why Are Speeders, Cheaters, and Bots Bad?

SPEEDERS

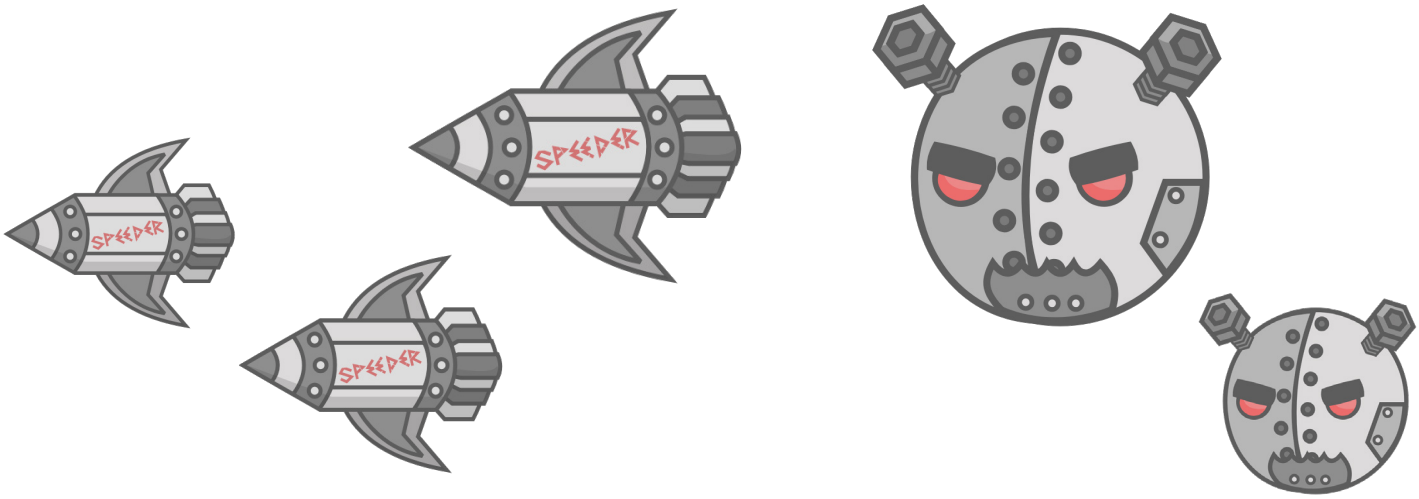
Speeders are very likely not reading or understanding your questions. This means they are not providing thoughtful responses and are just looking to collect their incentive as quick as possible. **Speeders are often "professional" respondents** who actively seek out surveys as a means of earning supplemental income.

CHEATERS

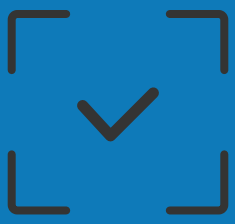
Cheaters may be reading and answering questions carefully, but not for the right reasons. Instead of providing honest answers they can provide answers they believe will get them into and/or through the survey quickly. In addition, cheaters may not even have any knowledge or experience with the subject matter, which could damage your data quality.

BOTS

Bots are bad because the data is **100% fake**. While some of the data collected from speeders or cheaters may be the actual opinions of a human, none of the data collected from a bot is. There are a surprising number of bot programs that exist, so you must design your survey instrument in a way to combat this.



Catching Speeders, Cheaters, and Bots



SPEEDERS

Survey speeders are one of the easier culprits to catch, as they somehow start and finish surveys at a rate that defies all the laws of physics. To catch them we recommend using a three-tier system based on the median time for completes.

TIER 1 NON-SPEEDERS

Those that finish the survey in a period longer than 50% of the median would pass as tier 1, non-speeders.

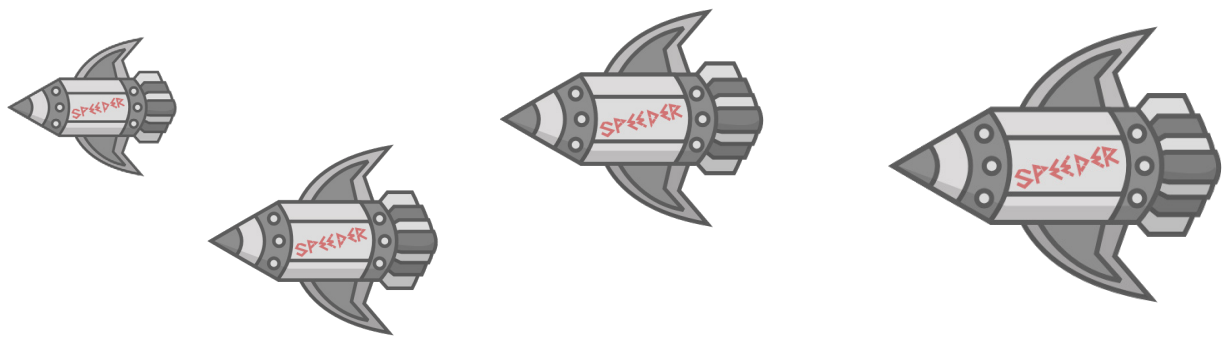
TIER 2

Those that finish in a period 50% of the median or shorter but longer than 30% of the median time would fall under Tier 2 and get a strike against them. (We'll talk more about this later.)

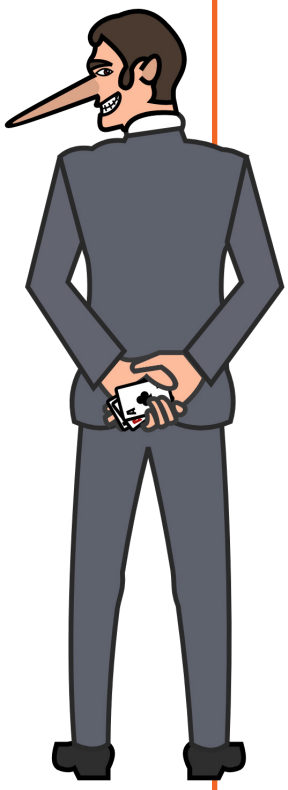
TIER 3 SPEEDERS

Those that finish in 30% of the median time or less and would just be thrown out altogether. The percentages vary from research team to research team, so you'll have to figure out what thresholds you are personally comfortable with.

**Keep in mind the average time to complete a survey may be different on a mobile device so creating separate thresholds based on device may be more accurate.*



CHEATERS



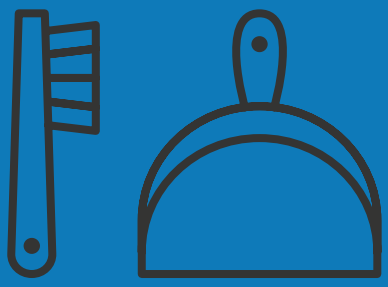
The best way to catch cheaters that try to qualify for your study is to **conceal the correct answers** to the screener. For instance, a question that asks, “Are you an anesthesiologist?” with a “yes” and “no” answer is generally only seen in a survey looking for anesthesiologists. A professional respondent will pick up on this right away and, if they are inclined to get into the survey, they might just decide that they are an anesthesiologist that day. Alternatively, a question that asks for the specialty of the respondent and has several answers, including the answer you are looking for, will be harder to guess their way into.

The inverse of this technique can also be used. You could, for instance, add a question that asks, “Are you a pediatrician?” when you are looking for anesthesiologists. Then you can toss those that say yes. Just be sure that it is very unlikely, or impossible for the respondent answer yes to the terminating question and be in the group of respondents you are looking for.

BOTS

The main way to thwart a bot is to limit its ability to learn. **Bots use a system of trial and error to get through surveys and learn more each time they try.** Make sure that your links do not allow this type of activity is key. The lesser the number of chances the bot has to try different paths, the less likely it will be successful. For surveys with a reward incentive, make sure the link is not an open link (able to be taken without an ID). If you have a list, add a unique ID and only allow those IDs into the system. Be sure that IDs are not easy to guess (i.e. incremental IDs). You can also check timestamps for batches of survey attempts within seconds of each other, which may indicate you have a bot on your hands.

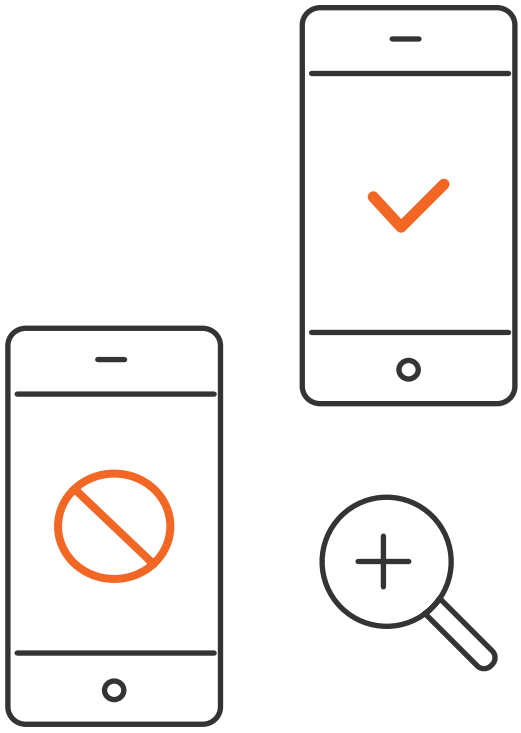
Lastly, ask your programmer if they have automated systems to check for bots and/or duplicate respondents. There are systems that can be used that give additional safety checks and data around real open-end answers, multiple attempts from the same machine, and other bot activity.

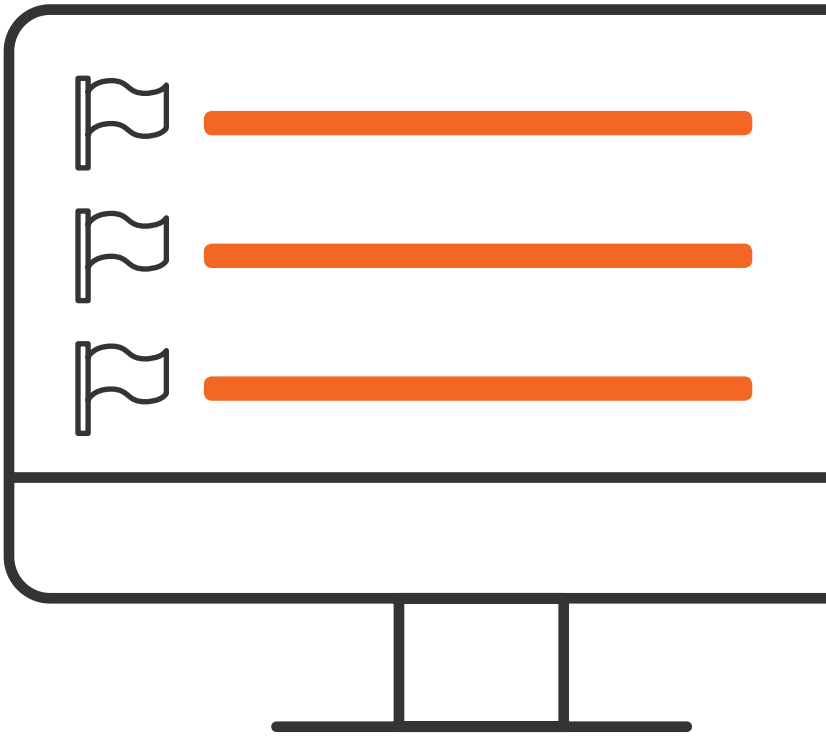


Cleaning Speeders, Cheaters, and Bots from Your Data

As you might have noticed there is some leeway around what makes a good respondent and a bad respondent. **While there are some activities that should disqualify respondents outright, there are others that may cause some concern, but are ultimately okay.** These items are good candidates for a “Strike” or Flag” system.

Adding and tracking flags throughout your survey can let you offer the leeway while giving you a way to check the respondents answers overall. Flags can be created manually in the data set or programmed into the survey itself to save time on the back end. Again, you can use tiers with flags as well. For instance, if you have 7 flags and a respondent triggers 1 or 2, you might not be concerned while you might kick out respondents with 5+ and further scrutinize the data for those with 3 or 4 strikes.





WHAT CAN YOU FLAG?

SPEEDING

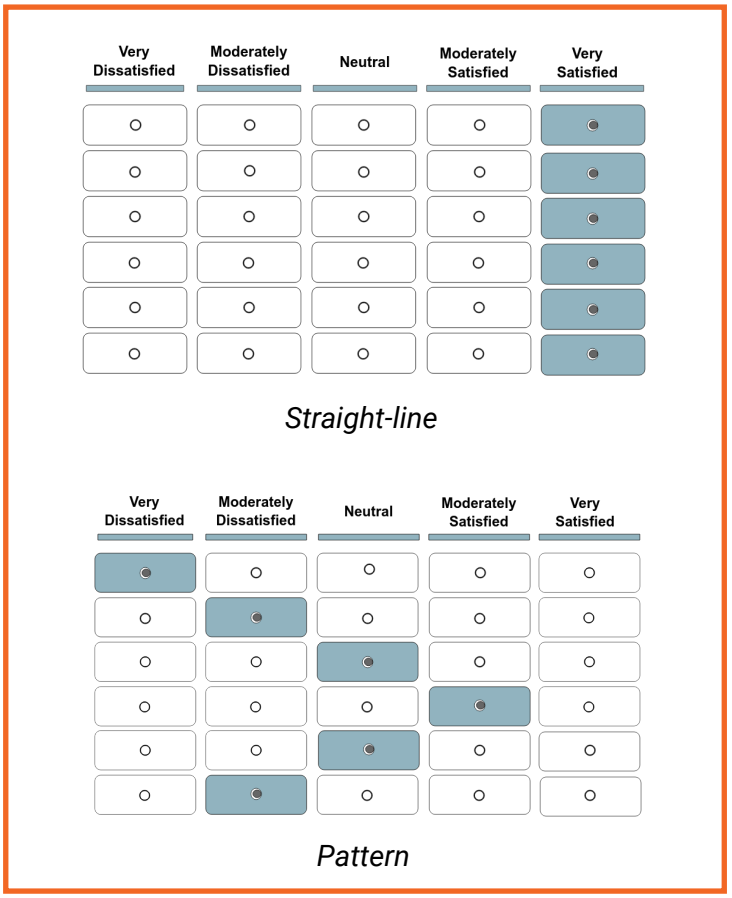
As previously mentioned, you can identify speeders based on a timeframe such as median time taken to complete the survey. Determine what an acceptable time range is for completing your survey and flag the respondents who don't fall within the range.

STRAIGHT-LINE/PATTERN RESPONSES

Speeders and cheaters can both try to get through a survey by answering questions in a straight line (answering the same for all items) or pattern.

The problem with blindly removing straight liners from your data is that there is a chance that they could really feel the way that they are answering. A more sophisticated way to check for pattern responses is to put two contradictory items in the answer group.

A straight-line series of answers in this example would be contradictory. As item 1 and 4 are close to opposites of each other.



Straight-line

Pattern

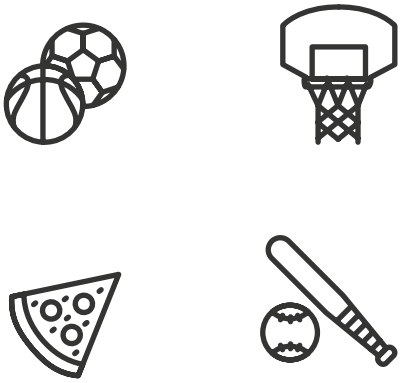
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
I often have trouble falling asleep at night	●	○	○	○	○
I exercise dailey	○	○	●	○	○
I never eat before bedtime	○	○	○	●	○
At bedtime I fall quickly and easily	●	○	○	○	○

FAILED TRAP/RED HERRING QUESTIONS

Use Trap/Red herring questions to catch disengaged respondents. These are questions that explicitly tell the respondent what to answer. For instance, “Please select ‘Very Satisfied’ here.” Red hearing questions can also be intentionally erroneous like a question of “Which of the following sports do you play? having an answer choice of “Pizza.” **Failing one trap/red herring question may not be enough to exclude a respondent from your data**, so you may want to sprinkle in a few and/or check to see if there are other suspicious answers.

Which of the following sports do you play?

- [x] Baseball
- [x] Basketball
- [x] Pizza
- [ ] Football
- [ ] Hockey
- ○ None



How satisfied or dissatisfied are you with company name in the following areas?

	Very Dissatisfied	Moderately Dissatisfied	Neutral	Moderately Satisfied	Very Satisfied
Price	○	○	○	◉	○
Quality of work	○	○	◉	○	○
Please select ‘Very Satisfied’ here	○	○	◉	○	○
Communication	○	○	○	◉	○

INCONSISTENT/CONTRADICTORY ANSWERS

Checking for inconsistent/contradictory answers is another method for detecting bad data. By putting in questions that allow for contradictory answers like “What is your age?” in the screener and “What year were you born?” in the demographic question. Creating questions that will catch inconsistent or contradictory answers can take additional planning and time however, they are a good way to **sniff out seasoned speeders and cheaters** who have caught on to other disqualifying behavior such as straight lining.

SUSPICIOUS/GIBBERISH OPEN-ENDS

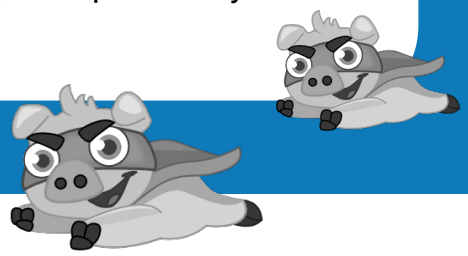
Open-ended responses can also help you catch speeders, cheaters, and bots. Check your data to look for signs of keyboard mashing, short, repeating, and nonsensical responses.

What do you think the main message of the ad is?

Eknfpwe nfipnwp	Yes
Gibberish/Keyboard Mashing	*Short/Nonsensical*

Preventing Speeders, Cheaters, and Bots

Although you can identify and remove survey speeders, cheaters, and bots from your data, the best-case scenario is to prevent them from getting into your data in the first place. **With a little planning you can reduce the amount of bad data you receive**, saving yourself from having to comb through countless numbers of respondents flagged as potentially bad data.



PREVENT DUPLICATES

Speeders, cheaters, and bots alike can be stopped by preventing respondents (or bots) from entering your survey more than once. If you are working with a specific list of respondents, the easiest way to do this is to create a unique ID for each survey link. If unique IDs are not an option, you use a digital fingerprinting to prevent respondents from taking your survey multiple times form the same machine.

WARN THEM

Build a flag tracking system into your survey that will warn respondents who are speeding or cheating that if they continue, they will be disqualified.
A simple reminder that you will only accept respondents who are carefully reading and answering each question may be enough to stop them in their tracks.

SCREEN OUT UNQUALIFIED RESPONDENTS

A well-designed screener is kryptonite for speeders, cheaters, and bots. Screeners are typically designed to target, and segment-specific respondents based on things like demographics and purchase behavior. However, with a little ingenuity, you can **design questions that will catch inconsistent or contradictory answers** and disqualify those respondents before they enter your main survey saving you from having to manually remove them from your data.

PREVENT SURVEY FATIGUE

There is a good chance that the respondent **did not** enter your survey with the intent to speed or cheat. They may have started out carefully reading and answers your questions but somewhere along the line starting speeding or cheating as a result of survey fatigue, which occurs when the respondents start to feel like your survey is **too long and/or boring.**

Reduce the Number of Questions

The more questions you have the more likely survey fatigue will set in. **Reducing the number of questions a respondent sees, does not necessarily mean you have to reduce the number of questions in your survey**. If you have a lot of questions you need answers to but want to keep your survey at a reasonable length you can implement branching logic, so respondents only answer questions that are relevant to them. Additionally, you can create multiple survey paths and split up respondents between them. For example, if you have 5 different product concepts, you could have each respondent only see one or two of them. Keep in mind doing this may require you to increase your sample size if you need the data to be statically significant.

Let Them Finish Later

Another way to prevent respondent fatigue is to give respondents the option to save their progress and finish taking the survey another time. If a respondent has the ability to finish the survey later, they may decide to take a break and come back later rather than speed through the rest of it. To make it as easy as possible you can give them the option to email themselves a link that will pick up from where they left off.

Make It Conversational

A survey should not feel like a test or read like a textbook. Your survey should be a dialogue between you and the respondent, and above all else, a good dialogue is about being conversational. Focus on writing questions that **adopt a casual tone** and then frame them in ways that replicate the way you'd talk to someone standing in front of you.

Make It Easy

Your questions and writing should be straightforward and the process of taking the survey should follow suit. If there are tedious, derivative questions, then you're going to lose their interest, damaging the honesty of their answers. A complex question may provide the best answers, but if no one reads the question, or gets confused while reading it, their answers, or lack thereof, could spoil your data.

Make It Interactive

Adding **interaction** to your survey will break up the monotony of a traditional survey by using multiple senses to create an engaging and even fun survey experience. If you can keep your respondents engaged, they will be less likely to get bored and abandon the survey. Interactive surveys can be created by using one or more of the following features:

Images / GIFS

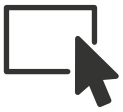
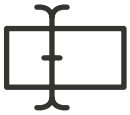


Image Clicking

Audio


Highlighting

Video

Swiping

Sliders

Gamification

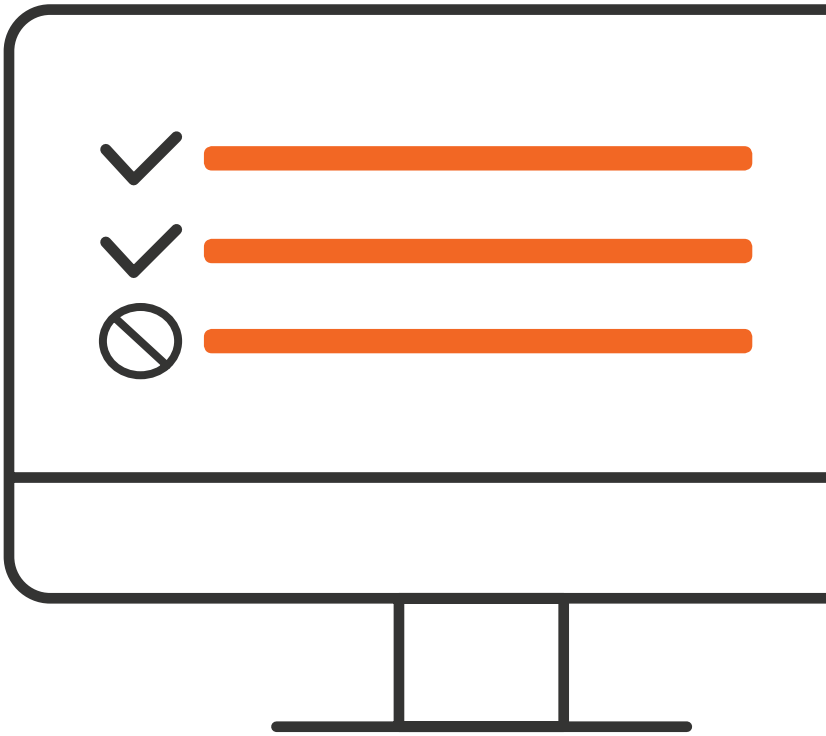
Drag & Drops

Custom Questions / Exercises

CONCLUSION

It is generally a good idea to remove respondents who have been flagged as bad data. However, it might be worth evaluating their responses to see if there is any viable data in there. Depending on the amount of bad data you received, they may not have a significant impact on the results. Consider looking at the data with and without flagged respondents to see how much damage they are doing.

In the end, **the process for catching and preventing speeders, cheaters, and bots is both a science and an art**. There is often trial and error while figuring out which thresholds work, what red herring questions are more effective, and which techniques work for you overall. Combining intelligent survey design with good survey system implementation will keep you on the road to clean survey data.



Jibunu is a pioneer of technology-enabled research services. Since 2003, we have been known for superior survey programming and research tool development.

WHAT IS JIBUNU?

The best in…

- Survey Programming
- Technology Consulting
- Custom Tool Creation
- Qualitative Tools
- Process Automation
- Collaboration
- Much More

WHAT WE DO

QUANTITATIVE: Complex logic and branching, trackers, segmentation, patient charts, concept testing, discrete choice, shopping exercises, custom-built tools, and much more!

QUALITATIVE: Interview flow programming, secure content hosting, interview scheduling, interactive exercises and process automation.

SUPPORT SERVICES: Consultation, process analysis and enhancement through technology, broadcasting, product photography, incentive fulfillment, coding, data analysis, sample procurement, and translation.

CONTACT US:

978-537-5510

EMAIL

JIBUNU.COM

310 AUTHORITY DRIVE, FITCHBURG, MA 01420