

More time for science: using Testable to create and share behavioural experiments faster, recruit better participants, and engage students in hands-on research

Constantin Rezlescu^{1,2}, Iulian Danaila², Alexandru Miron², Ciprian Amariei^{2,3}

¹Department of Experimental Psychology, University College London

²Testable

³Department of Computer Science, Al. I. Cuza University

ABSTRACT

A major pain for researchers in all fields is that they have less and less time for actual science activities: reading, thinking, coming up with new theories and hypotheses, testing, analysing data, writing. In psychology, three of the most time-consuming non-actual science activities are: learning how to program an experiment, recruiting participants, and preparing teaching materials. Testable (www.testable.org) provides a suite of academic tools to speed things up considerably. The **Testable software** allows the development of most psychology experiments in minutes, using a natural language form and a spreadsheet. Furthermore, any experiment can be easily converted into a social experiment in **Testable Arena**, with multiple participants interacting and viewing each other's responses. Experiments can then be published to **Testable Library**, a public repository for demonstration and sharing purposes. Participants can be recruited from **Testable Minds**, the subject pool with the most advanced participants verification system. Testable Minds employs multiple checks (such as face authentication) to ensure participants have accurate demographics (age, sex, location), are human, unique, and reliable. Finally, the **Testable Class** module can be used to teach hands-on psychology. It features over 50 ready-made classic psychology experiments, fully customisable, which instructors can add to their classes, together with their own experiments. These experiments can then be made available to students to do, import, modify, and use to collect data as part of their class. These Testable tools, backed up by a strong team of academic advisors and thousands of users, can save psychology researchers and other behavioural scientists valuable time for science.

INTRODUCTION

A major complaint of researchers in all fields is that they have less and less time for actual science. “Actual science” comprises of activities such as reading, thinking, coming up with new theories and hypotheses, testing, analysing data, writing. Instead, much of researchers’ time is spent on activities needed to support their science. In psychology, three of the most time-consuming supporting science activities are: learning how to program an experiment, recruiting participants, and preparing teaching materials. The first two problems also apply to commercial behavioural research. In fact, these issues are likely to be exacerbated in the industry because of generally tighter time frames for conducting a study. Because of the minimum technical skills required (i.e. working with a spreadsheet) and extremely fast learning curve (less than one hour), Testable really opens us the gates of experimental behavioural research to anyone interested in it.

Problem 1: Learning how to program an experiment

Experiments play a vital role in psychology. They allow researchers to test their hypotheses and advance our understanding of the human mind and behaviour. Whereas in early days experiments were conducted with pen and paper, nowadays most experiments are done on computers, requiring a new skill: knowing how to program. Universities have introduced year-long programming courses to all psychology degrees and students now spend a considerable amount of time learning various programming languages (often, more time than they spend studying cognitive psychology, or other branches of psychology). Programming frequently features among the most difficult topics for psychology students everywhere.

Note that this does not have to be this way. Mastering programming should not be a prerequisite for becoming an excellent psychology researcher. To get around this barrier, many departments and students/researchers use software designed to help create psychology experiments without coding (e.g. PsychoPy, OpenSesame). However, even these tools have a long learning curve to master their complex graphical user interfaces and then frequently require hours to create new experiments. The long time required to create experiments is a problem even for researchers who are expert programmers, who still need hours (sometimes days) to code an experiment from scratch.

Problem 2: Recruiting participants

Collecting data in the lab is a time-consuming process. It requires careful and extensive planning because, most often, participants cannot be run in parallel due to space, equipment, or experimenter limitations (i.e. not enough rooms, computers, or experimenters). To mitigate these shortcomings, more and more researchers choose to collect their data online. However, in practice, this comes with its problems, such as limited controls over the accuracy of the demographic data, whether the participants are humans or bots, participant engagement etc.

Problem 3: Preparing teaching materials

The best way to learn psychology is through experiments. However, re-creating classic psychology experiments (or creating novel experiments) for demonstration in class takes time and, as discussed above, programming knowledge from instructors. Furthermore, students learn best when they can modify play with the experiments, modify their parameters, collect data, and observe the effects of these changes on the results. If such experiments are created with a

programming language or a GUI software, the students need to invest substantial time to be able to make these changes.

Testable (www.testable.org) is a web platform offering a suite of tools for psychology research and teaching. These tools are: **Testable** (the software), **Testable Arena**, **Testable Library**, **Testable Minds**, and **Testable Class**. To use these tools, one needs to create a single Testable account (free when signing up with an academic email). Departmental licenses cover unlimited use of most features at no additional cost to individual researchers. Individual subscriptions or pay per use are also available.

Testable was kick-started with a small grant from the Association for Psychological Science awarded in 2015 (APS Fund for Teaching and Public Understanding of Psychological Science). We are also grateful for the financial support we have received from Ophelia Deroy and Bahador Bahrami (from an Arts & Humanities Research Council grant, Science in Culture, RTS-2013-2017), Brad Duchaine, Alfonso Caramazza, and Tirta Susilo.

TESTABLE - THE SOFTWARE

Testable makes it easy to create, run, and share behavioural experiments without any programming knowledge. It does not even require learning a complicated graphical user interface like other solutions do (e.g. one software boasts that their solution saves time by not having to learn how to program an experiment, while asking the user to learn no less than three different graphical user interfaces!). With Testable, researchers simply use a *natural language* form to communicate how their experiment should look like (see Figure 1 for a visual comparison of different solutions). Testable generates the experiment based on these specifications, and researchers then have the option to make finer adjustments on the spreadsheet associated with the experiment (*the trial file*, see below). With this method, creating a new experiment becomes extremely fast - most experiments will require only a few minutes from idea to being ready to run. Importantly, Testable is not designed to help only researchers without programming skills. Even experienced programmers may find it more convenient to use Testable than to write bespoke scripts to create their experiments.

Because it is so fast and easy to use, many new users wrongly assume Testable can be used only for simple behavioural experiments. In fact, Testable is comprehensive and very versatile, covering the vast majority of experimental paradigms. It can be used for all kinds of stimuli presentation (images, words, sounds, movies) and surveys. The response options include mouse clicks, button clicks, key presses, text input, voice, radio buttons, checkboxes, sliders, drag and drop and so on. Testable allows for almost unlimited randomisation options (including true randomisation), both within and between blocks, random sampling, random or counterbalanced allocation to subject groups, instant feedback based on responses, logic and branching. Researchers have full control over all the standard experimental parameters (e.g. presentation time, for simultaneous or sequential stimuli presentations, inter-trial interval, inter-stimulus interval, response window). A complete list of options is available in the manual available here: www.testable.org/manual.pdf. We detail a few of these options below, together with an explanation of how to use Testable to create your experiments and surveys. Note that Testable can be used online *and* offline because experiments are preloaded in the browser and can be run later without a working internet connection. That makes also performance independent of the quality of the internet connection.

A:

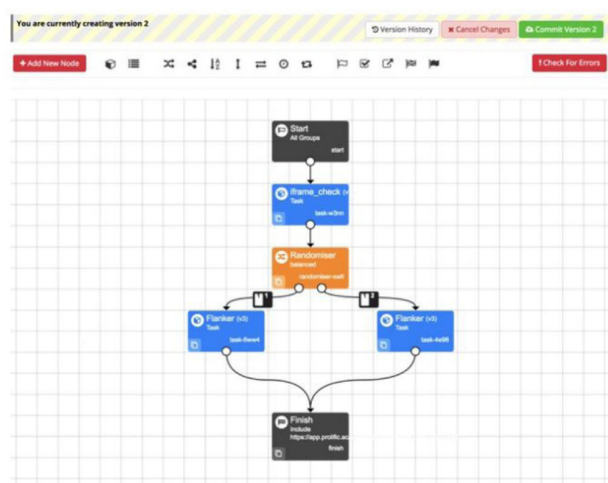
```
% load pictures & make textures .....
image=imread(stimulus1);
switch orientation
..... case 1
..... stim1=Screen('MakeTexture',D.window,image);
..... case 2
..... stim1=Screen('MakeTexture',D.window,image(end:-1:1,end:-1:1,:));
end
image=imread(stimulus2);
switch orientation
..... case 1
..... stim2=Screen('MakeTexture',D.window,image);
..... case 2
..... stim2=Screen('MakeTexture',D.window,image(end:-1:1,end:-1:1,:));
end
image=imread(mask);
msk=Screen('MakeTexture',D.window,image);
.....
% wait for the inter-trial interval to end
while GetSecs < (trial_start + DATA.ITI);
end
while KbCheck
end
end
```

B:



C:

The screenshot shows the 'Flanker' software interface. At the top, it says 'You are currently creating version 4'. Below this, there are tabs for 'Task Structure', 'Spreadsheet', 'Stimuli', 'Manipulations', and 'Script'. The 'Task Structure' tab is active, showing a tree view of the task structure. Under 'Instructions', there are 'Screen 1', 'Screen 2', 'Screen 3', and 'Screen 4'. Under 'Practice_Trials', there are 'Screen 1' and 'Screen 2'. The 'Screen 2' is expanded to show a 'Screen Layout' with a visual representation of the trial layout, including 'Horizontal Target' and 'Horizontal Distractors'. Below the layout, there are 'Configuration Settings' for 'Image' and 'Horizontal'.



D:

Tell us what you need

I want to present 2 images per trial, format jpg, displayed simultaneously for unlimited time.

I do not want to show a text above the stimuli.

Participants can respond by pressing keys: a,s,d. There is a correct response defined, and participants will not receive feedback after every trial.

I need 1 screen(s) with instructions, 10 practice trials and 20 main trials. Trials should not be randomised.

→ GENERATE PROJECT

Figure 1. A visual comparison of the different methods currently available to create a behavioural experiment: by programming (A), by learning how to use a graphical user interface, such as the one from PsychoPy (B) or three GUIs as required by Gorilla (C; note that only two are shown here), or by using the natural language form in Testable (D).

How does Testable work? Natural language and a trial file

Testable experiments are created according to the information provided in the *trial file*. The trial file is the backbone of any Testable experiment. The natural language form is just a shortcut to create this trial file. Researchers can choose to create the trial file from scratch without using the natural language form.

The trial file is essentially a CSV spreadsheet (Figure 2). Therefore, the most advanced technical skill a researcher would need for Testable is knowing how to fill in this spreadsheet.

The organisation of the trial file is as follows:

- 1) *Each row is a trial.* Note that the term “trial” is used in a broad sense; it refers to experimental trials, but also instructions screens. It also refers to questions in forms;
- 2) *Each column is a parameter defining the trials.*

type	stim1	stim2	stimFormat	presTime	title	content	button1	keyboard
instructions					Welcome!	Press 1 or 0	NEXT	
test	image1	image2	.jpg	2000				1 0
test	image1	image2	.jpg	2000				1 0

Figure 2. Example of a trial file. Each row (including the instructions) is a trial, and each column a parameter to define the trials. The first trial is used to display instructions, with a button labelled NEXT on which participants need to click to advance. The following two trials present two images, side by side, for 2000 ms. Participants have unlimited time to press 1 or 0 to advance to the next trial.

The columns in the trial file have suggestive names, e.g. stim1 for the first stimulus, stim2 for the second stimulus, presTime for presentation time. Researchers can add or delete columns/parameters as needed (except for the column “type” which is currently required for all

trials). A complete list of possible columns/parameters is provided in the manual, with columns organised by the function they support (e.g. columns to define and position the stimuli). Additionally, new columns may be added to help with organising the trials and the results. These columns will not be parameters used in the experiment, but will be saved in the results file, such that researchers can summarise results per condition, for example.

Testable projects appear in the main dashboard, with their current status, link, and areas to access the general setup parameters, trial file and results (see Figure 3). There are also buttons for privately sharing the project or posting it to Testable Library.



Figure 3. Testable projects appear in the main dashboard with their current status, the link where the experiment can be accessed, and areas for the general setup parameters, trial file and results. There are also buttons for duplicating and sharing the project, privately or by posting it to Testable Library, and for deletion.

Main features and capabilities

Testable experiments can contain different types of trials. Instructions trials are used to display instructions. Test trials are used for stimuli presentation and recording of participant responses. Practice trials are similar to test trials, only that participant responses are not saved. Learn trials are used for passive stimuli presentation (i.e. no participant responses). Finally, form trials are used to create questionnaires.

All kinds of stimuli can be included in Testable experiments: words, images, sounds, and videos. Column *stimFormat* is used to define their format. By default, stimuli are centered horizontally and vertically, but they can also be positioned manually or randomly using column *stimPos*. Their order on the screen can be randomised with column *stimOptions*. Whereas multiple audio/video stimuli are always presented in a sequence, multiple word/image stimuli can be presented simultaneously (side by side; by default) or sequentially (by defining interstimulus intervals in the column *ISI*). Unless a presentation time is specified in column *presTime*, word/image stimuli are displayed for unlimited time (until participant response). There is also an option to loop through the defined sequence of stimuli until participants response.

When it comes to specifying how participants can respond, Testable enables two types of responses by default. If there is a single stimulus on the screen, the default response is an input box. For multiple stimuli on the screen (necessarily words or images), the default response is to click on any of them. If other response types are needed, they must be specified in the trial file. For example, buttons can be added on the screen by using columns *button1*, *button2*, etc. Specific keys can be enabled with column *keyboard*. Other possibilities (pixel or area clicks, single or multiple input boxes and comment boxes, slider, labels for drag and drop, voice responses) are enabled using column *responseType* and customised using column *responseOptions*. All visual response types are displayed below the stimuli by default, but they can also be manually positioned anywhere on the screen by using column *responsePos*. Column *key* is used to specify the correct response in a specific trial.

Experiments can also incorporate trial by trial feedback (defined with column *feedback*). Feedback can be displayed regardless of participants response or only for certain responses, if participants responded correctly or incorrectly, or if they timed out. Feedback can incorporate the actual response from the current trial or from any of the previous trials, or a composite measure (e.g. mean, maximum, minimum) from a range of trials.

Several timing parameters can be defined in a Testable experiment: the stimuli presentation times (column *presTime*), the time window for participants responses (column *responseWindow*), the feedback duration (column *feedbackTime*), the inter-trial interval (column *ITI*) and the inter-stimulus intervals (column *ISI*). Random sampling from a predefined set or range of value is possible for most of these times.

In addition to these basic components of a behavioural experiment, there are a few other elements that can be added: fixation crosses, text to be displayed in each trial (typically with instructions above the stimuli), timer, counters for trial and block number, text boxes for the top left and top right corners of the screen.

Randomisation

Testable provides an extensive range of options for randomisation, both within a block (through column *random*) and between blocks (through column *randomBlock*). It is also straightforward. For example, any consecutive trials having the same number in column *random* will be randomized. Similarly, block randomization will be executed on blocks of trials as defined with *randomBlock*. Importantly, it is possible to define nested blocks with *randomBlock*, such that block randomization can occur at multiple levels and be as granular as needed.

By default, randomisation relies on the standard `Math.random()` JavaScript function (<https://tc39.es/ecma262/#sec-math.random>). This function offers random or pseudo-random numbers with approximately uniform distribution, based on browser implementation. For example, V8, the JavaScript engine used in the Chrome browser has `Math.random()` implemented using the xorshift128+ algorithm, a Pseudo-Random Number Generator (PRNG). In Testable, researchers can opt for “true” randomisation that uses a True Random Number Generator (TRNG). TRNGs use various physical phenomena as randomisation sources. Testable uses random.org services that are based on atmospheric noise.

Selection of trials: subject groups and random sampling

Some experiments require that participants see only a subset of trials defined in the trial file. Trials can be assigned to participants depending on their group (in between-subjects experiments) or randomly. In the first case, Testable uses column *subjectGroup* to mark the trials belonging to each group (e.g. *subjectGroup*=1). At the start of the experiment, participants are assigned randomly to a group and then only see trials from that group, plus the trials common to all groups (which have *subjectGroup* empty). The subject group allocation can be counterbalanced. In the second case, Testable uses column *randomPick* to randomly select a subset of trials to be presented to each participant. The size of the subset is the number used in *randomPick* to mark the consecutive trials from which the sampling should take place. For example, *randomPick*=20 will randomly select 20 trials from all the consecutive trials marked with *randomPick*=20.

Adaptive experiments

With Testable you can direct participants along multiple routes or branches in your experiment according to their responses. This is sometimes called experimental logic or branching. In Testable it is implemented with two columns: *if* and *then*. Column *if* specifies the condition(s) and column *then* specifies the action(s) to be taken when conditions are met (or not). Column *if* can refer to the participant’s response, whether it was correct or incorrect, and their reaction times, for the current or previous trials. Column *then* specifies the trial that participants should see next, when each condition is met, or when none of the conditions are met. The trial can be referenced

by the row number where it is found in the trial file or relative to the current point in the experiment (e.g. go back or forward two trials). Testable allows for multiple conditions with an action defined for each of them. Conditions can include negations and combine responses and reaction times.

Staircases

One particular type of an adaptive environment is the staircase procedure for finding sensitivity thresholds, originally used in psychophysics. In Testable, researchers can run staircases by defining all possible levels as consecutive trials, ordered from the easiest to the most difficult. To signal that these trials make up a staircase, they need to have the same label (e.g. A) in column *staircase*. Then, for the first trial in the staircase, the rules for the staircase should be defined: the starting point, the rule for increasing and decreasing difficulty (how many correct/incorrect responses, respectively), the steps for increasing/decreasing difficulty (how many levels to go down or up, respectively, in the trial sequence), and the rules for ending the staircases (number of reversals or number of trials). Multiple staircases can be combined into a group, such that participants “jump” randomly between trials belonging to different staircases (see Figure 4).

type	staircase	stim1	stimFormat
test	A: group 1, start 3, down 1, up 2, correct 3, incorrect 1, reversals 3	image1	.jpg
test	A: group 1	image2	.jpg
test	A: group 1	image3	.jpg
test	A: group 1	image4	.jpg
test	A: group 1	image5	.jpg
test	B: group 1, start 3, up 2, reversals 2	sound1	.mp3
test	B: group 1	sound1	.mp3
test	B: group 1	sound1	.mp3
test	B: group 1	sound1	.mp3
test	B: group 1	sound1	.mp3

Figure 4. Example trial file for two staircases (A and B) with five levels each, belonging to the same group 1. Participants would switch randomly between seeing an image from staircase A or hearing a sound from staircase B, depending on the level they reached in each staircase.

Setup options

Researchers have several other options available to them under setup. These options set the environment for each experiment, e.g. password protection, participant data collection, audience restrictions, custom debrief screen, results visualisation. Testable has created a number of ready-made start screens which can be added. There is a general instructions screen, a consent screen, a calibration screen to ensure all visual content is shown in the same size regardless of screen resolution and browser zoom level, and a screen to record participant details. Researchers can also select the language of the experiment (English, French, Dutch, Hebrew, Mandarin, Farsi currently available) and whether it should run in full-screen mode. Background and text colour are also customisable. Under the Audience options, researchers can set a maximum number of participants required and opt to restrict participation by IP address and country. Furthermore, researchers can customise the end screen, providing static or random completion codes and links for redirect.

Results files

Results can be saved in the cloud, locally, or both. For saving in the cloud, a working internet connection is necessary at the end of the experiment when results are sent to the server. Results are saved in one CSV file per participant, with information about the local machine (e.g. browser version, operating system, screen resolution), information about each trial, timing performance indicators (e.g. actual presentation time) and information related to participant actions: their response, reaction time, and whether the response was correct or not. Result files from multiple participants can be aggregated in a wide format spreadsheet, with each participant's responses in a row, ready to be imported into a statistical program such as SPSS or R for data analysis.

Timing performance

Timing performance refers to the accuracy (mean) and precision (variance) of: i) stimuli presentation time, and ii) reaction times. A recent study comparing various popular packages reported that web-based solutions, such as Testable, typically displayed an inter-trial variability of less than 5 ms for stimulus duration for a wide range of machine configurations (Bridges, Pitiot, Macaskill, & Peirce, 2020). While this was slightly larger than the results found with native desktop applications, it is in practice unlikely to affect the vast majority of psychology experiments. Note that even these relatively recent results from Testable are outdated. The testing was done in August 2019 (personal correspondence) and since then further optimisation changes related to timing performance were implemented in December 2019.

For reaction times, the reported inter-trial variability for browser-based experiments was under 10 ms in all cases in line with variability found for desktop software. These results were consistent with those found previously by de Leeuw & Motz (2016) who reported that, although response times in experiments running on JavaScript (in the browser) were approximately 25 ms longer than those found for experiments running on a desktop solution, there was no reliable difference in the variability of the RT distributions related to browser-based vs. non-browser-based experiments. Therefore, for behavioural studies relying on RT *differences* between experimental conditions (which, again, make up the vast majority of behavioural studies), experiments running in the browser are as sensitive as experiments running on desktop software. Consistent with this conclusion, another recent study looking at various standard psychology effects found no differences in data quality between the web-based and desktop-based experiments (Miller, Schmidt, Kirschbaum, & Enge, 2018).

Who is using Testable

There are now over 3,400 researchers and students who used Testable at least once to create an experiment. These users come from 100+ leading universities in US, UK, Germany, France, Switzerland, New Zealand, Australia, Iran, China. Additionally, several Psychology departments in UK, Switzerland and US acquired collective "umbrella" licenses.

There are already numerous peer-reviewed studies relying on Testable experiments. It is not possible to mention all here, but papers have been published in journals such as: PNAS (Phillips & Cushman, 2017), Psychological Science (Papeo, Stein, & Soto-Faraco, 2017), Scientific Reports (Smith, Bartholomew, Burnham, Tillmann, & Cirulli, 2017), Cerebral Cortex (Vannuscorps, F Wurm, Striem-Amit, & Caramazza, 2019), Cortex (Biotti et al., 2017; Ward & Filiz, 2020), Cognitive Psychology (Foster-Hanson & Rhodes, 2019), Evolution and Human Behavior (Jones, 2018), Neuropsychologia (Jiahui et al., 2017; Vannuscorps & Caramazza, 2016),

Journal of Experimental Psychology: Human Perception and Performance (Rezlescu, Susilo, Wilmer, & Caramazza, 2017), Attention, Perception & Psychophysics (Won & Geng, 2019), Experimental Brain Research (Forbes & Hamilton, 2017), Journal of Autism and Developmental Disorders (Sabatino DiCriscio & Troiani, 2017), Psychonomic Bulletin and Review (Devue & Grimshaw, 2018), Vision Research (Balas & Saville, 2017; Hacker, Meschke, & Biederman, 2019), Visual Cognition (Finzi, Susilo, Barton, & Duchaine, 2016), Perception (Kramer & Reynolds, 2018), PeerJ (Kramer, Mulgrew, & Reynolds, 2018), BMJ Open (Scherf et al., 2018), Personality and Social Psychology Bulletin (Rocklage & Fazio, 2018).

Testable is also a companion website for the fourth edition of Jamie Ward's popular textbook *The Student's Guide to Cognitive Neuroscience* (2020).

TESTABLE ARENA

Testable Arena is a multiplayer environment that can be enabled for any Testable experiment, effectively turning any Testable experiment into a social experiment. Therefore, the first step is to create a Testable experiment and then check the box for the multiplayer environment under the setup section. When this is done, a button "Start session" will appear next to the project name. Experimenters need to start a new session every time they want to run a multiplayer experiment. When a new session is started, experimenters are taken to the Monitor Room. Here they can see the multiplayer session link (which needs to be sent to participants to join) and the active participants waiting for the experiment to be started. At this stage, experimenters can rename participants, assign them to groups, or upload pictures to represent them. When ready, experimenters can start the experiment, after which moment no more changes are possible (and the session is closed for other participants). The Monitor Room will show all actions taken by participants during the experiment, in real-time, including responses and text messages (if chat was enabled).

Testable Arena consists of four possible spaces over-imposed on a regular Testable experiment: Participants space, Data Summary space, Chat space, and Log space. Most options related to these spaces are defined in the experiment setup, including which spaces to display (note that only three of these four spaces can be displayed simultaneously). There are also trial by trial options defined in the Trial File. The multiplayer environment is enabled at the trial level, such that an experiment can feature both single-player and multiplayer trials. Participants will complete some trials "alone" (i.e. without chatting with or seeing other participants' responses) and other trials in the "social" context offered by Testable Arena, in line with the experimenter's requirements specified in the trial file.

The Participants Space is for showing information about participants. Participants can be represented by genderless avatars, gender-specific avatars, or a webcam photo taken at the start of the experiment. Participants can also be identified by name, gender or age. Participant names/codes can also be assigned automatically. Images and names can be overwritten by the experimenter before the start of the multiplayer session. Additionally, it is also possible to display each participant's response in the current trial, and whether the response was correct or not.

Importantly, participants can be grouped. In this case, the groups will become the actors/agents of the experiment. A participant in a group will be able to see their in-group participants' information and behaviours, but only the aggregate behaviour of the other groups. From outside a group, it is not possible to know whether you are playing with a group or with a

single individual. A group can have a single individual. For example, you could have one group of four male participants and a second group of six female participants. The male participants will see each others' information and behaviour, but they will see only the aggregate behaviour of the female group (they will not even know they are dealing with a group). The number of groups must be pre-defined in the setup. The rule used for aggregating group responses is also defined in the Setup (the experimenter can select between: mean, modal, min, max, range, set, distribution).

The Data Summary space is dedicated to displaying participants' responses (or response times) in the current trial. Data can be shown by participants (i.e. what was the response of each participant), responses (i.e. which participants selected this vs. that response), or Correct/Incorrect (i.e. which participants responded correctly and which incorrectly). Participants (or groups) can be displayed with names or images.

The Log Space can be used to display responses (or response times) from all participants from all trials, whereas the Chat Space can be used to enable communication between participants before, during, or after a trial. The chat messages reset on every trial.

To mark which trials should be multiplayer, experimenters need to use column "arena" in the Trial File. This column can specify which Arena spaces (Participants/Data/Log/Chat) to show in each trial, or to go with the spaces defined in the setup. Three more columns ("arena_sequence", "arena_advance", "arena_responses") can be used to specify the sequence used to enable participants' responses (e.g. simultaneous or random sequential responses), to set the rule for moving to the next trial (e.g. automatically after all participants responded, or after a time delay), or to specify when to display participants' responses (e.g. instantly or at the end of a trial, after all participants responded).

Results for Arena experiments are saved in one large .csv file with all participants/groups, and one separate .csv file per participant. There is also a log saved as a .txt file (containing all events happening during the session, as seen in the Monitor Room, including chat messages with timestamps).

Testable Arena was developed with financial support from an Arts & Humanities Research Council grant (Science in Culture, RTS-2013-2017) to Ophelia Deroy, one of our academic advisors.

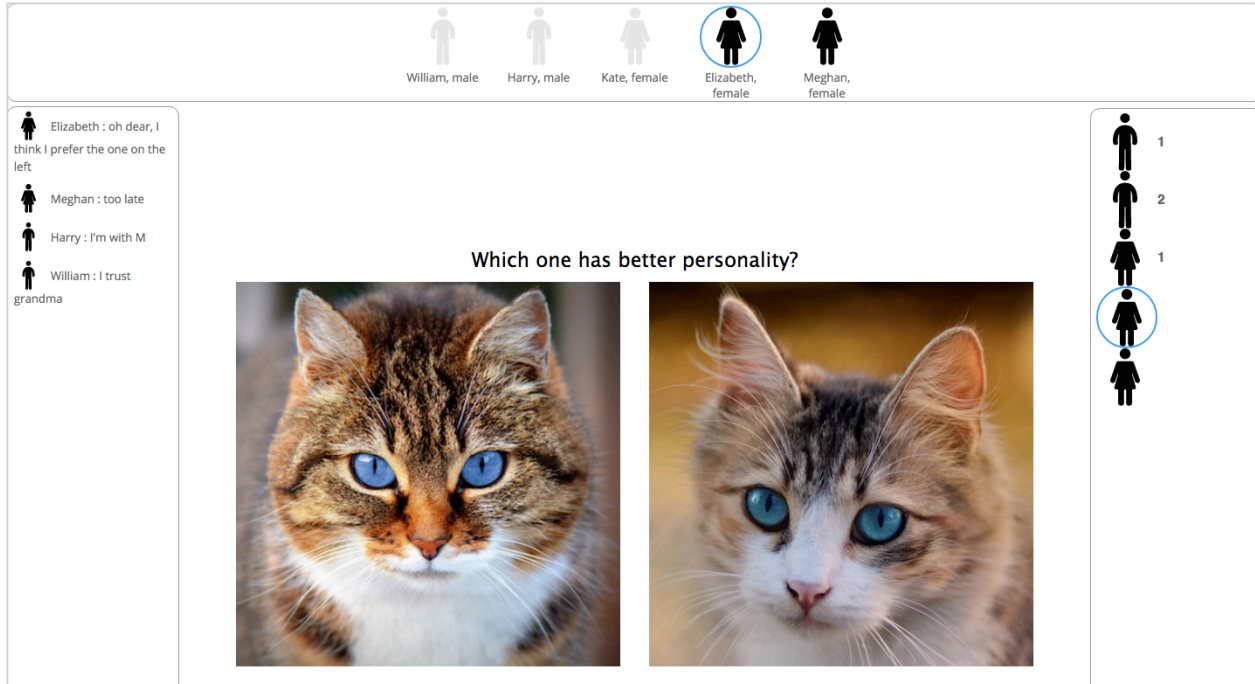


Figure 5. Illustration of a Testable Arena experiment, with five participants represented by avatars. Participants who responded are greyed out, current participant is marked with a blue circle. Participants can chat during the trial. The task is to select the preferred image.

TESTABLE LIBRARY

Testable Library (available at testable.org/library) is a repository of Testable experiments, created for two purposes: demonstration and sharing. With the current calls for more open science practices, it becomes important to offer researchers a space where they can make their published experiments publicly available to the wider academic community. No Methods section in a published paper, no matter how detailed, can compare to actually seeing and doing an experiment to understand it fully.

Any Testable experiment can be added to Testable Library. The workflow is straightforward. For the project you want to make public, click on the "Share" button, select "Send a copy to Testable Library", and fill in the requested fields with information about the name of the experiment, duration, and citation (if available). Important to know, this process will post an independent clone of your existing project to Testable Library. That means the changes you make to your project in Testable will not be reflected in the version available on Library. To update an experiment available on Library, you need to go through the same process and overwrite the existing experiment.

After posting, your experiment will be public at a permanent link (which you can add to the manuscripts describing your experiment). Anyone with the link will be able to run it. Other Testable users will also be able to request it from you, for use in their research. All requests must be individually approved by you before the experiment is shared with the requester. Of course, you can request experiments from other researchers too.

TESTABLE MINDS

Recruitment of online participants in psychology research has exploded in recent years. Platforms such as Amazon Mechanical Turk have facilitated the matching of researchers with large samples of willing participants and the collection of experimental data at a lower cost and in a fraction of the time normally required with lab samples. It is then not surprising that a recent review found that around 30% of articles published in three major journals for cognitive science mentioned the use of such platforms (Stewart, Chandler, & Paolacci, 2017). Importantly though, these platforms are essentially online labour markets, with participation in studies seen as a part-time job and participants looking to maximise their winnings through participation in as many experiments as possible (Berg, 2016). Half of the surveyed MTurkers reported participation in more than 160 academic studies in a month (Kees, Berry, Burton, & Sheehan, 2017). The professionalisation of participation in psychology experiments can negatively affect the validity of data collected, with newer results being influenced by past participations (Chandler, Mueller, & Paolacci, 2014).

Lower confidence in data quality from typical labour market participant pools

Because of the potential for more and more earnings depending on their participation in as many studies as possible, participants in typical labour market pools such as MTurk and Prolific may be tempted to game the system. In fact, researchers recently reported a flurry of random responses in such studies (Bai, 2018), which may or may not have been produced by bots. To our knowledge, none of these pools employ comprehensive filters to restrict access from the most likely locations for bot-like or random responses: IP anonymizing services, VPNs (Virtual Private Networks), server hosting providers, public proxies and Tor exit nodes. Furthermore, there are very limited efforts to verify participant identity and basic demographic information.

What distinguishes Testable Minds from other participant pools

Testable Minds is a participant pool created to avoid many of the issues affecting existing solutions to recruit online participants (Hauser, Paolacci, & Chandler, 2019). First of all, it is exactly that: a participant pool for psychology experiments, not a labour market. Our recruitment efforts target individuals who show an interest to help behavioural research without having a financial reward as the main motivator. Furthermore, when new participants sign up to become a “testable mind”, they are explicitly told that Testable Minds is not a money-making platform, is not a labor market, and that their winnings (and thus participation) will be limited to \$50 per month.

Second, and probably most importantly, Testable Minds employs the most advanced participant verification system. The goal is to guarantee that all participants are: *Human, Unique, Verified* for basic information, and *Reliable*. *Human* means ensuring that it is not possible (nor profitable) to use bots to complete experiments. *Unique* means making sure the same individual cannot create multiple accounts with Testable Minds, nor participate multiple times in the same experiment. *Verified* refers to making sure everyone is who they say they are, by confirming basic information such as age, sex, and location. *Reliable* refers to the quality of participation – a human, unique and verified participant can still provide poor data quality (because of a lack of attention, for example). Testable Minds employs verification checks at sign up and for each study participation. Some checks occur only at sign up or during participation, while others are performed at both (e.g. reCaptcha verification against bots).

Verification at sign up

After passing a commonly used verification against bots (reCaptcha v2, Google), participants are required to provide: i) an email, Google, or Facebook account; ii) a phone number; iii) a Revolut or PayPal account for money withdrawals. All these elements contribute to verifying uniqueness. The email account is confirmed by asking participants to click on a confirmation link sent to that email account, and the phone number is confirmed by asking participants for a confirmation code sent in a text message to that number. Additionally, the phone number is confirmed only if the country code matches the country of the IP address from which Testable Minds is accessed. We use an external solution to screen out unreliable locations (see information on MaxMind service below). Participants who pass these tests are then asked to fill in basic information: first and last names, date of birth, sex at birth, nationality, and country of residence.

To be qualified as a “verified mind”, our participants are then required to submit an official photo ID (passport, national ID, or driving license) and take a live webcam photo of their face. The photo ID is used to manually verify the first and last names, date of birth, sex, and country of residence (and also the nationality, depending on the ID) by a Testable team member. Then, we use the FaceRekognition algorithm from Amazon to perform a face matching between the photo on the ID document and the webcam photo. This result is also confirmed by visual inspection. Only after all these tests are passed successfully, the participant is qualified as a “verified mind”.

Face authentication and other verification for each study participation

For each study participation, regardless of the platform used to host the study (Testable or other), participants must pass stringent verification tests. These are set in place to prevent automated software (bots) or scammers start the experiment: i) reCaptcha v2 by Google, and ii) MaxMind GeoIP2 Precision Insights Service [<https://www.maxmind.com/en/geoip2-precision-insights>]. MaxMind service allows Testable Minds to filter any access from an anonymizing service, VPN (Virtual Private Network), server hosting provider, a public proxy or Tor exit node.

For “verified minds” an extra step is required to start an experiment: *face authentication*. Participants are asked to take another live webcam photo, which must match the one saved at registration (which in turn matched the one on the official photo ID). This process is automatically completed using Amazon Rekognition – CompareFaces service [<https://docs.aws.amazon.com/rekognition/latest/dg/APICompareFaces.html>].

If the study requires participants from specific locations (researchers can select these when posting a study), the current participant location is identified based on the user’s IP address using the MaxMind GeoIP2 Country Database. The service offers a 99.8% accuracy [<https://www.maxmind.com/en/geoip2-country-database>]. This verification adds another level of confidence regarding location at the time of participation, beyond the phone and country verification at registration.

“Minds” can participate only once in a study.

Size and demographics

As of January 2020, there were 11,400 “minds” registered in our subject pool, 51.7% male and 48.3% female, with a mean age of 34 years old (SD = 11.5). In terms of location, 42% were from US, 19% from UK, 18% from EU countries, 2% from Canada, and 2% from Australia and New Zealand.

How to post a new study

Testable Minds can be accessed from the main menu by any researcher with a Testable account. To post a new study for data collection, researchers need to fill in a form with basic information about the study, such as the title, short description, area, estimated duration, whether the study includes audio/video stimuli or long written assignments. For experiments featuring graphic images or strong language, researchers are required to upload their approved ethics application (in practice, researchers upload their approved ethics for most studies, even standard ones). Researchers can add several selection criteria, such as location, age, sex, first language, and past approval rate. It is also possible to create minds panels, with participants in previous studies, and post the new study only for those participants (e.g. for longitudinal studies). Finally, researchers must fill in the participation fee (minimum hourly rate is \$6) and the number of participants required. After the necessary budget (participation fees plus commission) is covered, researchers can send the study for approval. All studies for Testable Minds are reviewed before being posted for data collection.

Note that any browser-based experiment can be posted to Testable Minds, not only those created with the Testable software. For the non-Testable experiments, researchers are required to provide a completion code to participants.

TESTABLE CLASS

Testable Class is a module specifically created for instructors who want to incorporate experimental demonstrations in their teaching and engage students in hands-on research. Classes are essentially collections of experiments created by an instructor, for which students can sign up. Instructors can create new experiments, import existing ones from their Testable account, or select from a list of ready-made classic psychology experiments developed by Testable. This list was created taking into account the most common psychology experiments and phenomena described in psychology textbooks. Once added to the instructor's class, these tests are fully customisable.

Within a class, instructors can choose when to make experiments visible to students, which ones to be completed for credit, and whether to allow students to import and modify them (Figure 6). If students are allowed to import but not modify an experiment, all results that they collect will appear under the instructor's account as well. For example, such an experiment shared with a class of 20 students each collecting 20 participants will appear with 400 results under the instructor's account. Allowing the students to also modify an experiment enables them to experiment with various parameters and learn first-hand about the effects of these modifications on the results. Instructors see a summary of credits, number of experiments and recorded results for all students.

Testable Class has been effusively received in liberal arts colleges from US (e.g. Wellesley College) where there is a strong emphasis on teaching.

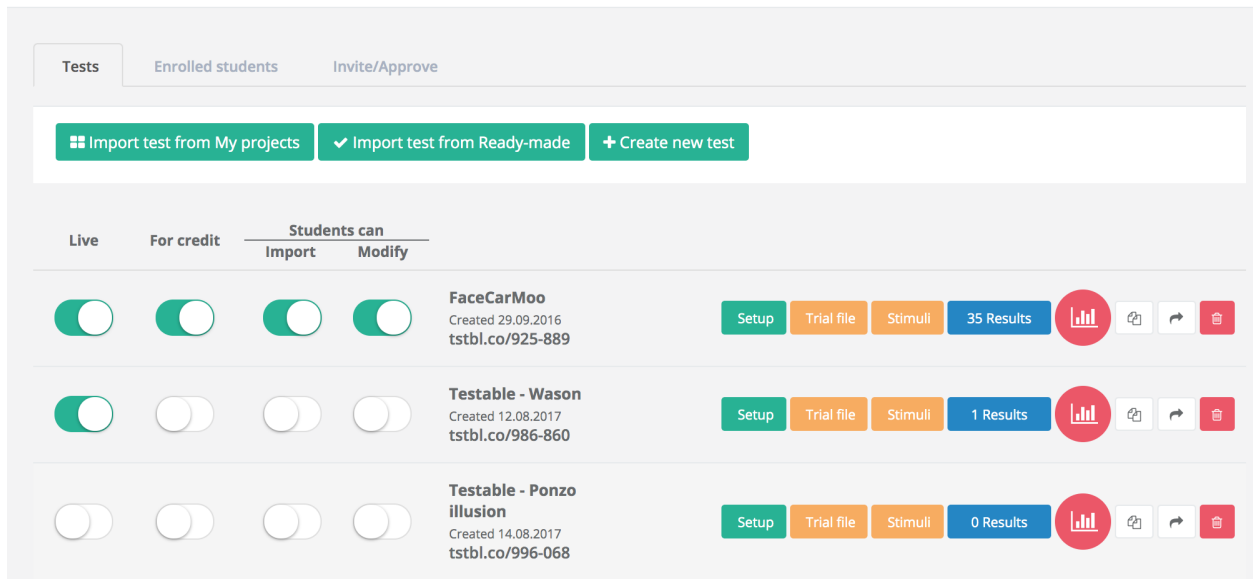


Figure 6. Example of a class in Testable Class. Tests can be added by importing existing Testable projects, creating new ones, or importing from the ready-made collection of common psychology experiments. Researchers can decide when to make tests available (live) to students, which ones to mark for credit, and which ones to allow students to import and/or modify.

CONCLUSION

The web-based platform Testable contains a comprehensive suite of advanced, yet easy to use tools to optimise behavioural research and related teaching. The software **Testable** eliminates all technical barriers to create state-of-the-art behavioural experiments, from the simple to the most complex ones, such as staircases, adaptive and social (interactive) experiments, dramatically reducing the time between idea and testing. **Testable Minds** proposes the most advanced participant verification system for online studies (including face authentication), ensuring that participants are human, unique, and verified for basic demographics. **Testable Class** is used by instructors to teach psychology through ready-made and customised experiments. We encourage behavioural scientists, in academia and industry, to give them a try. Testable will continue to evolve based on user feedback and to innovate along with technology advancements, under the guidance of our strong team of academic advisors (<https://www.testable.org/#advisors>).

- Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research*, *141*, 228–236. <https://doi.org/10.1016/j.visres.2016.12.005>
- Berg, J. (2016). Income security in the on-demand economy: findings and policy lessons from a survey of crowdworkers. *Comparative Labor Law and Policy Journal*, *37*(3), 506–543.
- Biotti, F., Wu, E., Yang, H., Jiahui, G., Duchaine, B., & Cook, R. (2017). Normal composite face effects in developmental prosopagnosia. *Cortex*, *95*, 63–76. <https://doi.org/10.1016/j.cortex.2017.07.018>
- Bridges, D., Pitiot, A., Macaskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PsyArXiv*. <https://doi.org/10.31234/osf.io/d6nu5>
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130. <https://doi.org/10.3758/s13428-013-0365-7>
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. <https://doi.org/10.3758/s13428-015-0567-2>
- Devue, C., & Grimshaw, G. M. (2018). Face processing skills predict faithfulness of portraits drawn by novices. *Psychonomic Bulletin and Review*, *25*(6), 2208–2214. <https://doi.org/10.3758/s13423-018-1435-8>
- Finzi, R. D., Susilo, T., Barton, J. J. S., & Duchaine, B. C. (2016). The role of holistic face processing in acquired prosopagnosia: evidence from the composite face effect. *Visual Cognition*, *24*(4), 304–320. <https://doi.org/10.1080/13506285.2016.1261976>
- Forbes, P. A. G., & Hamilton, A. F. d. C. (2017). Moving higher and higher: imitators' movements are sensitive to observed trajectories regardless of action rationality. *Experimental Brain Research*, *235*(9), 2741–2753. <https://doi.org/10.1007/s00221-017-5006-4>
- Foster-Hanson, E., & Rhodes, M. (2019). Is the most representative skunk the average or the stinkiest? developmental changes in representations of biological categories. *Cognitive Psychology*, *110*(December 2018), 1–15. <https://doi.org/10.1016/j.cogpsych.2018.12.004>
- Hacker, C. M., Meschke, E. X., & Biederman, I. (2019). A face in a (temporal) crowd. *Vision Research*, *157*, 55–60. <https://doi.org/10.1016/j.visres.2018.02.007>
- Hauser, D., Paolacci, G., & Chandler, J. (2019). Common concerns with MTurk as a participant pool: Evidence and solutions. In F. Kardes, P. Herr, & N. Schwarz (Eds.), *Handbook of Research Methods in Consumer Psychology*. <https://doi.org/10.4324/9781351137713>
- Jiahui, G., Garrido, L., Liu, R. R., Susilo, T., Barton, J. J. S., & Duchaine, B. (2017). Normal voice processing after posterior superior temporal sulcus lesion. *Neuropsychologia*, *105*, 215–222. <https://doi.org/10.1016/j.neuropsychologia.2017.03.008>
- Jones, A. L. (2018). The influence of shape and colour cue classes on facial health perception. *Evolution and Human Behavior*, *39*(1), 19–29. <https://doi.org/10.1016/j.evolhumbehav.2017.09.005>
- Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155. <https://doi.org/10.1080/00913367.2016.1269304>
- Kramer, R. S. S., Mulgrew, J., & Reynolds, M. G. (2018). Unfamiliar face matching with photographs of infants and children. *PeerJ*, *2018*(6), 1–25. <https://doi.org/10.7717/peerj.5010>

- Kramer, R. S. S., & Reynolds, M. G. (2018). Unfamiliar Face Matching With Frontal and Profile Views. *Perception*, 47(4), 414–431. <https://doi.org/10.1177/0301006618756809>
- Miller, R., Schmidt, K., Kirschbaum, C., & Enge, S. (2018). Comparability, stability, and reliability of internet-based mental chronometry in domestic and laboratory settings. *Behavior Research Methods*, 50(4), 1345–1358. <https://doi.org/10.3758/s13428-018-1036-5>
- Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, 28(3), 369–379. <https://doi.org/10.1177/0956797616685769>
- Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences of the United States of America*, 114(18), 4649–4654. <https://doi.org/10.1073/pnas.1619717114>
- Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 43(12), 1961–1973. <https://doi.org/10.1037/xhp0000400>
- Rocklage, M. D., & Fazio, R. H. (2018). Attitude Accessibility as a Function of Emotionality. *Personality and Social Psychology Bulletin*, 44(4), 508–520. <https://doi.org/10.1177/0146167217743762>
- Sabatino DiCriscio, A., & Troiani, V. (2017). Autism-like Traits are Associated With Enhanced Ability to Disembed Visual Forms. *Journal of Autism and Developmental Disorders*, 47(5), 1568–1576. <https://doi.org/10.1007/s10803-017-3053-0>
- Scherf, K. S., Griffin, J. W., Judy, B., Whyte, E. M., Geier, C. F., Elbich, D., & Smyth, J. M. (2018). Improving sensitivity to eye gaze cues in autism using serious game technology: Study protocol for a phase I randomised controlled trial. *BMJ Open*, 8(9). <https://doi.org/10.1136/bmjopen-2018-023682>
- Smith, L. M., Bartholomew, A. J., Burnham, L. E., Tillmann, B., & Cirulli, E. T. (2017). Factors affecting pitch discrimination performance in a cohort of extensively phenotyped healthy volunteers. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-16526-8>
- Stewart, N., Chandler, J., & Paolacci, G. (2017, October 1). Crowdsourcing Samples in Cognitive Science. *Trends in Cognitive Sciences*. Elsevier Ltd. <https://doi.org/10.1016/j.tics.2017.06.007>
- Vannuscorps, G., & Caramazza, A. (2016). The origin of the biomechanical bias in apparent body movement perception. *Neuropsychologia*, 89, 281–286.
- Vannuscorps, G., F Wurm, M., Striem-Amit, E., & Caramazza, A. (2019). Large-Scale Organization of the Hand Action Observation Network in Individuals Born Without Hands. *Cerebral Cortex*, 29(8), 3434–3444. <https://doi.org/10.1093/cercor/bhy212>
- Ward, J., & Filiz, G. (2020). Synaesthesia is linked to a distinctive and heritable cognitive profile. *Cortex*.
- Won, B.-Y., & Geng, J. J. (2019). Flexible target templates improve visual search accuracy for faces depicting emotion. *Attention, Perception & Psychophysics*. <https://doi.org/10.3758/s13414-019-01965-4>