

# YourSuper Comparison Tool

**Results from a survey and two survey experiments**

**October 2022**

**Other uses**

Enquiries regarding this license and any other use of this document are welcome at:

Managing Director

Behavioural Economics Team of the Australian Government

Department of the Prime Minister and Cabinet

1 National Circuit

Barton ACT 2600

Email: [beta@pmc.gov.au](mailto:beta@pmc.gov.au)

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Prime Minister and Cabinet or the Australian Government.

**Research team**

Australian Taxation Office Behavioural Insights Unit staff who contributed to this report were: Georgia Conduit, Indrani Das, and Luzian Ong. BETA staff (current and former) who contributed to the report were: Harry Greenwell, Laura Bennetts Kneebone, Daniel Carr, Michael Ridgway, Jack Hodge and Michael Bleasdale.

**Acknowledgments**

Thank you to the Treasury for their support and valuable contribution in making this project happen. In particular, thanks to Ben Dolman, Alex Maevsky, Bianca Bauer, and Katarina Trinh for their work on this project. Thanks also to ATO’s Design and Digital experience teams.

The trial was pre-registered on the American Economic Association registry:  
<https://www.socialscienceregistry.org/trials/6995>

Who?

### Who are we?

We are the Behavioural Economics Team of the Australian Government, or BETA. We are the Australian Government’s first central unit applying behavioural economics to improve public policy, programs and processes.

We use behavioural economics, science and psychology to improve policy outcomes. Our mission is to advance the wellbeing of Australians through the application and rigorous evaluation of behavioural insights to public policy and administration.

### What is behavioural economics?

Economics has traditionally assumed people always make decisions in their best interests. Behavioural economics challenges this view by providing a more realistic model of human behaviour. It recognises we are systematically biased (for example, we tend to satisfy our present self rather than planning for the future) and can make decisions that conflict with our own interests.

### What are behavioural insights and how are they useful for policy design?

Behavioural insights apply behavioural economics concepts to the real world by drawing on empirically-tested results. These new tools can inform the design of government interventions to improve the welfare of citizens.

Rather than expect citizens to be optimal decision makers, drawing on behavioural insights ensures policy makers will design policies that go with the grain of human behaviour. For example, citizens may struggle to make choices in their own best interests, such as saving more money. Policy makers can apply behavioural insights that preserve freedom, but encourage a different choice – by helping citizens to set a plan to save regularly.

Contents

[Executive summary 4](#_Toc115187658)

[Why? 6](#_Toc115187659)

[What we did 9](#_Toc115187660)

[Results 13](#_Toc115187661)

[Discussion and conclusion 20](#_Toc115187662)

[Appendix 1: Evaluation design and analysis 22](#_Toc115187663)

[Appendix 2: Other survey results 28](#_Toc115187664)

[Appendix 3: Statistical tables for survey experiments 32](#_Toc115187665)

[Appendix 4: Survey wording 42](#_Toc115187666)

[References 48](#_Toc115187667)

## Executive summary

### Why we undertook this project

Many Australians have invested their savings in superannuation funds that significantly underperform benchmark net returns. This is typically due to a combination of weak returns and high fees. Ultimately, underperformance can lead to retirement savings that are far lower than they might have been.

The *YourSuper Comparison Tool* website(‘the comparison tool’) was developed to empower superannuation fund members. The comparison tool allows people to compare the performance and fees of MySuper products, with the aim of making it easier for them to choose a well‑performing product that meets their needs.

### What we did

The Australian Taxation Office (ATO) led the design and development of the comparison tool, in collaboration with a multi-agency team comprising staff from the Treasury, the Australian Securities and Investment Commission (ASIC), the Australian Prudential Regulation Authority (APRA), and the Behavioural Economics Team of the Australian Government (BETA).

BETA and the ATO’s Behavioural Insights Unit worked together to contribute behavioural insights to the design process. This included running a survey and 2 survey experiments to generate relevant evidence. The survey, completed by 2,236 eligible respondents in January 2021, included questions on: when respondents would be likely to use the comparison tool, the best way to describe a poor performer, and, in which format they would prefer an ‘underperformance notice’ (mail or email).

### Results and limitations

In the first survey experiment, we tested how super products were classified and sorted on the results page of the comparison tool. They were either placed into 2 categories (‘poor’ or ‘good’ performance) or 3 categories (‘poor’, ‘fair’, or ‘top’ performance). As we expected, respondents who saw the additional ‘top’ performance category were better at selecting one of the best performing funds than those who saw only 2 performance categories.

In the second experiment, we tested different ways to communicate the risk/return trade‑off associated with different super products. For example we tested accurate but technical language (such as ‘low risk and low expected return’) against terms that are common in the super industry but where the risk level is implied (such as ‘conservative’, ‘balanced’ and ‘growth’). Respondents who were presented with common industry terms for risk/return were better at selecting the optimal option in the scenario presented.

Our overall judgement is that these results are likely to reflect how people would respond to these design options in the real world. Some caution is warranted since survey experiments only test responses in an artificial setting. Furthermore, online survey panels are not fully representative of the general population. However, we collected a large and diverse sample of superannuation members, and we designed an experiment that sought to mimic real‑world decision‑making.

This research was considered in the development of the YourSuper comparison tool alongside user research conducted by the ATO and consultation with the super fund industry and other stakeholders.

This report is accompanied by a spreadsheet appendix with full survey results (tabulated). It is available at: <https://behaviouraleconomics.pmc.gov.au/projects/yoursuper-comparison-tool-results-survey-and-two-survey-experiments>

## Why?

### Policy context

Australia’s retirement income policy is based on 3 pillars: compulsory superannuation, the age pension, and voluntary savings (Australian Government the Treasury 2020:1). The Government’s superannuation guarantee currently requires employers to contribute 10.5% of employees’ salaries into a superannuation fund, thereby reducing retirees’ reliance on the age pension. The decisions super fund members make about their superannuation throughout their lives can have a significant impact on their incomes in retirement.

Super fund performance significantly impacts their members’ retirement savings. For example, the Productivity Commission’s (PC) Inquiry Report *Superannuation: Assessing Efficiency and Competitiveness* estimated that being in an underperforming fund can lead to 54% less in retirement savings compared to being in a high‑performing fund (PC 2018a:11). Yet the PC identified many ‘entrenched underperformers’ (that is, super funds that consistently underperform relative to benchmarks) and many consumers remain in underperforming super products for long periods of time.

Choosing an appropriate super fund is challenging. There are many super products available, and several complex variables to consider, including the product’s past performance, fees, investment strategy, insurance policies and ethical investment options. Added to this, people lead busy lives and often don’t have time to devote to monitoring the performance of their super fund. Many people struggle to understand financial matters in general, with around 30% of Australians having low financial literacy (PC 2018a:21). In these conditions, many consumers are likely to find the choice overwhelming (Chernev et al, 2015). Often inertia will set in and they will make no ‘active’ choice at all, remaining in the same fund, with the same investment strategy for many years.

‘MySuper’ is the name given to default super products. That is, when employees start a new job, their super can be paid into a super fund of their choice or into the MySuper product chosen by the employer. If an employee doesn’t nominate any super fund (and doesn’t have an existing, eligible fund) then the employer *must* pay the employee’s super into a MySuper product.

MySuper is designed to simplify the choice of super product for new entrants to the Australian workforce and thus reduce the impact of choice overload and inertia. To qualify for MySuper, a super fund must meet extra requirements set by APRA. These products typically offer lower fees and simple features.

MySuper products are now a significant part of Australia’s super system. As of March 2022, approximately 27% of all super assets were invested in these products, in around 14 million member accounts (APRA 2022, and APRA 2021:table 5). However, even within this highly regulated segment of the market, the PC found that more than one million member accounts were in MySuper products that were underperforming (PC 2018a:13).

Finding relevant, accurate performance metrics for super funds can be difficult and time consuming. This makes it difficult for consumers to assess how a super product compares with the rest of the market, and which products might be better. In 2019, the Australian Prudential Regulation Authority (APRA) began publishing performance assessments of MySuper products via a product heatmap, as part of its supervisory focus on ensuring trustees are delivering appropriate outcomes for their members. APRA’s MySuper product heatmap was mainly targeted at super funds, rather than consumers.

To address the underperformance of MySuper funds, the *Your Future, Your Super* reform package included 2 initiatives.

* Funds with underperforming MySuper products (based on one of APRA’s performance indicators contained in the MySuper product heatmap) would be required to write to their members, notifying them that their product was underperforming. The aim of this initiative was to hold super funds to account for underperformance, protect super members from poor outcomes, and encourage funds to lower their costs and fees to boost Australians’ retirement incomes.
* A comparison tool was created to allow consumers to compare the performance of MySuper products. The comparison tool aims to empower super members by making it easier for them to choose a well‑performing product that meets their needs.

### The problem

It is not easy to design a comparison tool that is easy to use, and helps people engage with their superannuation. Seemingly small details can have outsized impact on whether users feel confident to choose a super product that is suitable for them.

There are many government and non‑government comparison sites for services such as energy, mobile and internet, insurance, and home loans. Most have the same basic features: a series of questions to identify what the user is looking for, a summary results page, more detailed pages for individual products, and instructions on switching to a new product.

The YourSuper comparison tool has a similar structure to other comparison sites, and includes some additional features. Importantly, it makes use of the outcomes of the legislated performance test. For users who login to myGov, it can be pre‑populated with a user’s super fund details, reducing the number of questions users must answer. Finally, the tool directs users to other services and information on superannuation offered by the ATO.

Our research focused on 2 issues relating to the design of the summary results page.

First, the policy intent was that the tool would indicate whether the super product was ‘performing’ or ‘underperforming’, based on the legislated performance test outcomes. However, this would still leave users to choose from a long list of ‘performing’ funds—a formidable task. Furthermore, investing in a high performing fund (compared to an average one) really matters for retirement savings due to the nature of compounding returns over long periods of time. We wondered whether the choice process could be improved by adding a third performance category (‘top’, ‘fair’ or ‘poor’).

Second, superannuation consumers also need to decide on the appropriate investment strategy for their super, and the corresponding trade‑off between risk and expected returns. It is important that this decision is presented in a way that people understand and allows them to make a choice consistent with their life stage and preferences. Selection of a sub‑optimal risk/return level can have very large impacts on retirement savings. Most notably, a young person who chose a low‑risk investment option would have much less at retirement than one who chose an option with a higher risk but higher expected return.

We were concerned that people may misinterpret common terminology around risk since it can be an emotive term. Specifically, consumers may not think about risk in terms of volatility of their investment portfolio, but instead in terms of the chance that they ‘lose their money’. The latter interpretation could lead people to make decisions that are unsuitable for their life stage (especially for young people). So, in our second survey experiment, we tested the effect of different risk terminology on the choice of investment strategy in a particular scenario.

## What we did

We conducted a survey from 7 to 14 January 2021. It was completed by 2,236 eligible respondents who were close to nationally representative in terms of gender, age, and metropolitan/regional split. The survey contained 2 survey experiments to test the impact of the presentation of super product performance and risk in the comparison tool. In addition, we included survey questions to investigate other aspects of engagement with super such as people’s preferences on communication formats and features of the comparison tool.

### Survey experiment one – presentation of performance categories

The YourSuper comparison tool is available on the ATO website. Users are given the option to access a ‘personalised version’ by logging in through myGov. In this case, key details (their age, current super balance, and super product details) are all pre‑filled. Alternatively, users can access a ‘non‑personalised’ version of the tool, which produces generic results. These results can be customised if the user enters their age and current super balance, or searches for their current super product.

The main results page of the comparison tool displays a list of super products along with their annual fee and their 7-year net return. In addition, each super product is placed in a performance category. In the first experiment, we varied the number of performance categories (2 or 3), and whether the results were sorted. We tested the impact of these variations on which super product was chosen by survey respondents.

We asked participants to respond to a hypothetical scenario based on a mock‑up of the results page of the comparison tool. The mock‑up was based on the ‘non‑personalised’ version of the tool, however, the scenario indicated that respondents had already customised their results by entering their age and super balance.

Survey experiment one – scenario presented to respondents

*Imagine you’ve been working for about a decade – you’re 32 years old. You’re reading a news article on your laptop that mentions a new government website – a superannuation comparison tool. The article also says that you can change your super fund if you want to. You follow a link, and go to the site. It prompts you to enter a few details, which you do. You put in your age (32), your current super fund (Stability Super), and your best estimate of how much super you currently have ($75,000). Click through to the next page to see your personalised results.*

Respondents were randomly assigned to see one of 3 different versions of the results table. The different versions varied in relation to the number of performance categories – either 2 tiers (poor, good) or 3 tiers (poor, fair, top). In addition, the sorting of results varied: products were presented either in a random order or sorted by performance category (with a random order within each performance category). This led to participants being sorted into three groups:

* Group A: 2 performance categories with results randomly sorted (figure 1)
* Group B: 3 performance categories with results randomly sorted (figure 2)
* Group C: 3 performance categories with results sorted by performance category (figure 3).

For simplicity, all tables were filtered so the results page showed products with a single risk/return level (‘Medium‑High’). The respondents were not able to interact with the table to sort the products, or check for further details on each product.

Respondents were asked to select which product they thought was the best option in the hypothetical scenario and we used this to construct 2 outcome measures:

* First, we assigned a ranking based on the product’s net return, with the top‑ranked product given a value of 10.
* Second, we looked at how many respondents chose one of the top 3 products.

We expected that respondents who saw 3 performance categories, and products sorted by performance category, would be more likely to choose the higher performing funds.

We also collected several secondary outcome measures including the time people took to complete the task, and respondents’ answers to questions about:

* which factor was most important in their decision
* how likely it is that they would switch super products
* how easy it was to compare products

1. Results presented to Group A, with 2 performance categories randomly sorted

**This image is of the mock results table that people in Group A saw. It shows 10 super funds and lists each funds annual fees, net return, risk level, and performance. 
There are blue text and arrows that were not on the table that respondents saw. They are there to highight key features of the Group A table for people reading the report. For Group A they point out that there were only 2 performance categories, and that the table is not sorted.**

1. Results presented to Group A, with 3 performance categories randomly sorted

**This image is of the mock results table that people in Group B saw. It shows 10 super funds and lists each funds annual fees, net return, risk level, and performance. 
There are blue text and arrows that were not on the table that respondents saw. They are there to highight key features of the Group B table for people reading the report. For Group B they point out that there were 3 performance categories, and that the table is not sorted.**

1. Results presented to Group A, with 3 performance categories sorted by performance

**This image is of the mock results table that people in Group C saw. It shows 10 super funds and lists each funds annual fees, net return, risk level, and performance. 
There are blue text and arrows that were not on the table that respondents saw. They are there to highight key features of the Group C table for people reading the report. For Group C they point out that there were 3 performance categories, and that the table is sorted by performance category.**

### Survey experiment two – risk terminology

In the second experiment, we explored how to communicate the risk/return trade‑off inherent in different investment strategies. We designed an experiment that tested whether the terminology used to describe risk would have an impact on the choice of investment strategy.

Respondents were asked to select the most appropriate superannuation investment strategy for a young person starting out in their career. In general, young people are best served by selecting a high risk/high return investment strategy. Our outcome measure looked at how many respondents selected this optimal option.

Respondents were presented with 3 investment risk options to choose from (covering low, medium and high risk‑and‑return settings). Each randomly assigned experimental group saw different descriptive labels for the same underlying investment strategies, as follows:

* **Group X Technical definition highlighting negative returns**: A direct statement of the risk level and a definition of what this means (e.g. ‘Low risk – expect a negative return 1 out of every 20 years’)
* **Group Y Common industry terms**: The investment strategy terms commonly used in the superannuation sector to convey risk (‘Conservative’, ‘Balanced’, ‘Growth’)
* **Group Z Technical definition highlighting expected returns**: The risk level and expected return (e.g. ‘Low risk & low expected return’)

Survey experiment two – question and response options for Group X Technical definition highlighting negative returns

*Which is the best option for a young person starting out in their career?*

* *Low risk – expect a negative return 1 out of every 20 years*
* *Medium risk – expect a negative return 2 out of every 20 years*
* *High risk – expect a negative return 4 out of every 20 years*

### Survey questions

The survey questions aimed to collect respondents stated preferences on issues around communication and information presentation. It included questions on the following topics;

* When people would be most likely to use the comparison tool
* Key data and useability features they are looking for in a super comparison tool
* The best ways to remind them to check the performance of their super fund(s)
* Which slogans, encouraging them to check the performance of their super, they found most motivating
* Their preferred communication channel for being notified that their super fund is underperforming
* Which wording best conveys a super fund’s performance.

A full copy of the survey questions is available in Appendix 4.

## Results

The results for the 2 survey experiments are set out below, followed by the survey findings.

### Survey experiment one – presentation of performance categories

We found that people who saw 3-tier performance categories in the results table were much more likely to select higher performing products. It appears that adding the ‘top’ performance category helped many people to narrow down their options before making a choice based on fees. Sorting the results table by performance categories had additional benefits.

#### Number of performance categories

We tested the impact of showing 2 or 3 performance categories by ranking the 10 products based on their net return (that is, their performance). We then calculated the average rank for each group based on the products selected by respondents in that group. We also looked at the proportion who chose any of the top 3 products.

Respondents who saw 3 performance categories (rather than 2) chose higher ranking products and were more likely to choose a ‘top’ product (one of the top three products by net return). For respondents who only saw 2 performance categories only 34% chose a top product. For respondents who saw 3 performance categories, 55% chose a top product, and this increased to 61% when categories were sorted (see Figure 4). These differences were statistically significant.

1. Respondents were more likely to select a product with a high net return when presented with 3 performance categories and a sorted table.

n=2,236. Note: a ‘top’ product was defined the top 3 options based on net return (secondary outcome measure)..

All groups could see the precise net return for all products, so what drove this increase in the preference for top performing funds? The proportion of people who selected a product from the ‘poor’ category was similar across all 3 groups.[[1]](#footnote-2) But people were much more likely to select a product with a high net return when it was actually labelled ‘top performing’. The ‘poor’ category was equally effective across the groups (as a deterrent) but, as expected, adding the ‘top’ category helped respondents choose between the remaining products.

We also asked respondents which element they were most focussed on in making their choice of product in the scenario. The most common response was ‘best performance’ (33%), then ‘highest returns’ (26%), and then ‘lowest fees’ (20%). By contrast, when we examined the choices that respondents actually made, the most popular in all three trial arms was the *lowest‑fee* product within the highest performance category. Notably, this was not the product with the highest net return.[[2]](#footnote-3) This suggests the ‘top’ performance category assisted many people to narrow down their options before making a choice based on fees.

#### Sorting the products by performance category

Sorting the super products by performance category provided additional benefits. Groups B and C both saw three performance categories but only in Group C were products sorted by these categories. Respondents in Group C had somewhat better results than Group B with an average rank of 7.0 vs 6.8 and 61% versus 55% choosing a top product. Both differences were statistically significant.

Notably, the results do *not* suggest a ‘donkey vote’ effect from sorting (choosing the option highest in the list without making a genuine decision). In all groups the first listed product was “Your current fund” as this was pinned to the top.

#### Intention to switch

Once respondents had chosen the fund they thought was best, we asked them how likely it was (in the hypothetical scenario) that they would switch to this fund. The three different versions of the comparison tool did not produce any practical differences in people’s *intention to switch*. For all 3 treatments groups the average response was between 3.5 and 3.6 (which represents somewhere between ‘neither likely nor unlikely’ and ‘likely’ on the Likert scale). It seems the changes in presentation helped people comprehend the information but did not increase the likelihood that they would report an intention to take action.This may be a function of the artificial, hypothetical task they were given, or it may reflect the need for additional interventions to boost the motivation to switch.

#### Ease of decision making

We were also interested in whether differences in the presentation of the results table made the decision‑making process easier. We looked at this by comparing the time taken to complete the decision task, and we also asked respondents to rate the difficulty of the task (on a 5‑point scale, ranging from ‘very easy’ to ‘very difficult’). On both measures, the outcome was similar across the 3 groups. For time taken to complete the task, the means of all 3 groups’ were between 50 and 53 seconds. For the task difficulty, all 3 groups’ means were around 4 (which represents ‘easy’). While both measures have limitations, this suggests that even if the presentation of the results table could improve the quality of respondents’ choices, we did not find evidence that it also made the choice process easier.

### Survey experiment two – risk terminology

We found that people were much more likely to select the optimal investment strategy for a young person when this strategy was described using common language used by super funds (that is, ‘Conservative, ‘Balanced’ or ‘Growth’).

Respondents who saw the common super industry terms – ‘Conservative’, ‘Balanced’, ‘Growth’ – were much more likely to select ‘Growth’, the most appropriate option for a young person (see Figure 5). 49% of respondents in this group selected a ‘Growth’ type investment strategy, compared to 27% for the group who saw technical definitions highlighting negative returns, and 30% for the group who saw technical definitions, highlighting expected returns. In the group who saw the common industry terms, the risk level was implied, whereas the other 2 groups both used a more technically accurate description that mentioned risk level explicitly.

1. Respondents were more likely to select the appropriate investment risk for a young person starting off in their career if it was described with a common industry term.

n=2,236. Note: the bars depict proportions for the 3 different groups so they are not required sum to 100%.

This suggests that directly referring to the risk level may deter individuals from selecting high‑risk investment options even when these may be the most appropriate to their age and circumstances. Surprisingly, this appears true even when the risk level is accompanied by a statement of the higher expected return (as in Group Z: Technical, highlighting expected returns).

### Survey Findings

#### People preferred performance categories to be labelled strong or poor performing

Respondents were asked to select the performance labels they thought would be most helpful in describing the best and worst performing super products. We were particularly interested in which terms would be meaningful for people with low ‘financial literacy’ so we disaggregated results by this variable.[[3]](#footnote-4) However, the following results should be interpreted with caution as they may have been influenced by the labels displayed in the first survey experiment, which used the label ‘poor’ for all participants, and ‘top’ for 2 of the 3 groups of participants.

For the best performing products, a larger proportion preferred ‘strong’ performing (39%) as opposed to ‘top’ (33%) or ‘high’ (27%). These proportions did not vary greatly by financial literacy level.

For the worst performing products, more respondents preferred ‘poor’ performing (44%) over ‘weak’ or ‘under’ (Figure 6). However the results for respondents with lower financial literacy were different: a somewhat greater proportion preferred ‘under’ or ‘weak’ (35-36%) over ‘poor’ (30%). This made it difficult to draw a firm conclusion about the best terminology for poor performing funds.

1. Preferred terminology for the worst performing super products was “poor performing”

Q Which label would be most helpful in highlighting the worst performing super funds? n=2,236.

#### People think they would use a super comparison tool when they receive their annual super statement

Respondents said they would be most likely to use a comparison tool to compare super products when they receive their annual super statement (39%, see Figure 7).

The strength of this preference differed by financial literacy level: 48% of high financial literacy respondents selected this option relative to 25% of those with low financial literacy although it remained the most popular option for both high and low financial literacy groups. The next most popular option for those with low financial literacy was if they heard about it on traditional media (23%, vs only 14 % for those with high financial literacy).

1. People were most likely to use a comparison tool when they get their annual statement.

Q. When do you think you would be most likely to use a superannuation comparison tool? Select all that apply. n=2,236.

#### People prefer to receive an underperformance notice via email

#### We asked respondents how they would like to be notified if their fund was performing badly (an ‘underperformance notice’). Overall, respondents displayed a clear preference for receiving the underperformance notice by email or both mail and email, rather than mail only (Figure 8).

1. People wanted to receive an underperformance notification by email

Q. You are provided with a notice telling you that your superannuation product is performing poorly and directing you to the superannuation comparison tool. Which delivery method would be most likely to get your attention? n=2,236.

#### Other survey results

We present additional survey results in Appendix 3 and we have provided a full tabulation of the results from all survey questions in the spreadsheet appendix.[[4]](#footnote-5) Briefly, the other main results from the survey were:

* The most popular slogan to motivate people to check their super performance (28%) stated: ‘*A small change now can make a big difference in retirement’*
* The most important features for a superannuation comparison site are that it be free (56% respondents stated this was ‘very important’) and easy to use (52%).
* Respondents indicated that a letter from their super fund (39%) would be the best type of communication to prompt them to check the performance of their super fund.

### Limitations

Our survey experiments were stylised simplifications of the actual experience of using a comparison tool. How respondents engaged with a static, stylised snapshot of one screen may not reflect how they would engage with an actual experience of using an interactive web‑based comparison tool. This simplified environment may cause our estimates of impact to be too high, or too low. In addition, we asked respondents about a hypothetical scenario where they were 32 years old, regardless of their actual age.

Our outcome measure assumed that one of the ‘top’ products would always be a more appropriate product. In reality, choosing a super fund is complicated and any comparison tool will necessarily simplify some variables (such as fees or returns) or omit others (such as insurance). Consequently, it is important to be mindful that the ‘top’ products will usually – but not always – represent an appropriate product, depending on an individual’s circumstances.

The sample drawn from an online survey panel may not be representative of all Australian superannuation fund members since people who sign-up for a survey panel may be systematically different from all superannuation fund members.

Online surveys can have data quality issues if, for example, respondents do not engage meaningfully with the questions (although we did take steps to manage data quality).

These limitations are discussed further in Appendix 1.

## Discussion and conclusion

Providing people with accurate super fund metrics, collated on a website from a trusted provider, is a useful contribution to improving outcomes for superannuation fund members. Our experiments provide evidence that the way the comparison tool presents information matters. Most notably, we found that presenting fund performance in 3-tier performance category groups (versus 2-tiered) led to people being more likely to select one of the higher net return super products. Sorting the order of the results by fund performance provided another boost. We also found that communicating risk using more direct, technical language (such as ‘low risk/low expected return’) may lead to poorer decision‑making.

A key tenet of behavioural insights is to make things easy. This project demonstrates the value in this approach. In a stylised survey experiment, small changes designed to simplify the comparison tool’s results table led to substantial changes in the super funds chosen.

The use of a third (‘top’) performance category was particularly effective. It is not easy to navigate a lengthy list of super funds’ performance metrics and work out which products have performed well. The extra performance category may assist people by offering an indication of what constitutes high performance. In addition, people may use the performance categories to narrow down their search (and simplify their decision) before deciding on their fund of choice.

We also found that ordering the results by performance mattered, but to a lesser extent. However our study may understate the impact of ordering since the actual comparison tool features more than 70 funds as opposed to the 10 in our study. In reality, the table of results stretches over multiple pages and so ordering the results by performance on the real world comparison tool is likely to have a greater impact than it did in our study.

Our project also explored the language used to describe the risk/return trade‑off in choosing different investment strategies. Broadly, our results suggest the language used in the comparison tool should be given careful consideration. It shouldn’t be assumed that the technical language used commonly in the literature or in policy circles is a better way to communicate to the general public (even if it is accurate), when compared to terms commonly used in the super industry. In particular, people may have a different understanding of, or response to, the term ‘risk’ than what is intended.

Our survey findings also suggest that ‘underperformance’ may not be the clearest term to use to describe a fund that has performed well below its benchmarks.

This study, like any, has limitations (described under ‘What we did’). The survey experiments were only stylised representations of real‑world decision‑making. In addition, the online survey panel may not be representative of the population of interest and could have generated some data quality issues, despite our efforts to manage these. However, we collected a large and diverse sample of superannuation members, and designed an experiment that sought to mimic real‑world decision‑making. Consequently, our judgement is that the central findings of this study are likely to reflect how people would respond in the real world to the design options we examined.

## Appendix 1: Evaluation design and analysis

### Overview

This project was a collaboration between BETA and the ATO Behavioural Insights Unit. We conducted an online survey and 2 online survey experiment in January 2021, using the online survey provider Dynata to source respondents.

In the first survey experiment we showed respondents a mock-up of the summary results page of the comparison tool and asked them to select a super product. Respondents were randomly assigned to see different versions of the summary results page. Key outcomes were super product chosen, as well as time taken to make a choice.

In the second experiment, we asked people to respond to a question about the appropriate risk‑return level for a young person starting out in their career. Respondents were randomly assigned to see different terminology to describe the risk-return level.

### Pre-registration, pre-analysis plan and ethics

We pre‑registered the trials, along with our pre‑analysis plan, on the American Economic Association RCT Registry (13/01/2021): <https://www.socialscienceregistry.org/trials/6995>. We registered the pre-analysis plan after the trial commenced but prior to accessing any of outcome data.

We did make one deviation from the pre‑analysis plan. For Experiment one, we did not run a robustness check that retained responses that took less than 2 minutes and 30 seconds to complete the survey.

The research was subject to ethics approval from Macquarie University’s Human Research Ethics Committee (reference number 52020944424157).

### Trial design

We conducted a survey between 7 and 14 January 2021. It was completed by 2,236 eligible respondents who were close to nationally representative in terms of gender, age, and metropolitan/regional split. The survey contained 2 survey experiments to test the variations to the summary results page of a comparison tool. It also included survey questions relating to various preferences for the comparison tool.

For each experiment, participants were randomised to a trial-arm within the survey platform, Qualtrics, with roughly equal allocation to each condition. The allocation to a given treatment arm in Experiment one was independent of the allocation to any arm in Experiment two.

### Experimental Outcomes

This section how the outcome variables were constructed for each experiment, consistent with the pre‑analysis plan.

#### Experiment one – outcomes

**Primary outcomes**

* + - 1. *Choice of fund*: This is an ordinal variable based on the answer provided when choosing a super fund from the list. Each fund was given a ranking ranging from 1-10, with 1 being the worst performing fund (lowest net return), and 10 being the best performing fund (highest net return). For the analysis this was treated as a continuous variable.
      2. *Time to make a decision*: This is a continuous variable based on how long participants spent on the experiment page (measured in seconds). We removed anybody who was more than 3 standard deviations above the mean. This has the effect of excluding implausible response times which occur when the respondent pauses the survey and completes it at another time.

**Secondary outcomes**

*Choice of ‘Top’ fund*: This is a binary variable indicating whether respondents chose one of the three funds with a performance rating of ‘Top’.

*Choice of any fund that is not ‘poor’ performing*: This is a binary variable indicating whether respondents chose a fund with either a ‘Top’, ‘Fair’ or ‘Good’ performance rating.

*Ease Rating*: This is an ordinal variable measured through a 1-5 Likert scale rating of how easy it was to complete the task (where 1 = very difficult, 5 = very easy). For the analysis this was treated as a continuous variable.

*Intention to switch*: This is an ordinal variable measured through a 1-5 Likert scale rating of respondents’ likelihood to switch funds if faced with that same scenario (where 1 = very unlikely, 5 = very likely). For the analysis this was treated as a continuous variable.

#### Experiment two – outcomes

*Primary outcome –* *Selection of correct investment risk answer*: This is a binary variable indicating if the respondent selected the most appropriate investment risk option for a young person starting their career (i.e. the higher risk, higher reward option).

### Interventions and hypotheses

This section describes the interventions (treatment arms) for each experiment, along with the hypotheses that were detailed in our pre‑analysis plan.

#### Experiment one – interventions and hypotheses

**Interventions**

Experiment one was a 3-arm trial. Participants saw a summary results page that varied in 2 ways: the number of ‘performance categories’ used to classify the super products, and how the super products were sorted. The 3 arms were as follows:

* *Group A*: two categories (Good and Poor performance), unsorted
* *Group B*: three categories (Top, Fair and Poor performance), unsorted
* *Group C*: three categories, sorted by performance category (then randomly sorted within each category).

**Hypotheses**

(All one-sided tests)

*Hypothesis 1.1*: On average, respondents in Group B (three categories, unsorted) will choose a higher rated fund as compared with Group A (two categories, unsorted) (Group A < Group B).

*Hypothesis 1.2*: On average, respondents in Group C (three categories, sorted) will choose a higher rated fund as compared with Group B (three categories, unsorted) (Group B < Group C).

*Hypothesis 1.3*: On average, respondents in Group C (three categories, sorted) will choose a higher rated fund as compared with Group A (two categories, unsorted) (Group A < Group C).

*Hypothesis 1.4*: On average, respondents in Group B (three categories, unsorted) will take less time to make a decision than Group A (two categories, unsorted) (Group A > Group B).

*Hypothesis 1.5*: On average, respondents in Group C (three categories, sorted) will take less time to make a decision than Group B (three categories, unsorted) (Group B > Group C).

*Hypothesis 1.6*: On average, respondents in Group C (three categories, sorted) will take less time to make a decision than Group A (two categories, unsorted) (Group A > Group C).

#### Experiment two – interventions and hypotheses

**Interventions**

Experiment two tested whether the selection of the most appropriate investment risk option varied by the terms used to describe the investment risk profile. It included the following groups:

* *Group X*: Technical definition, highlighting negative returns. (e.g. ‘Low risk – expect a negative return 1 out of every 20 years’)
* *Group Y*: Common industry terms. (‘Conservative’, ‘Balanced’, ‘Growth’)
* *Group Z*: Technical definition, highlighting expected returns. (e.g. ‘Low risk & low expected return’)

**Hypotheses**

(All 2‑sided tests)

*Hypothesis 2.1*: On average, respondents in Group Y (common industry terms) will select the most appropriate response at a different rate to those in Group X (Technical, highlighting negative returns) (Group X ≠ Group Y).

*Hypothesis 2.2*: On average, respondents in Group Z (Technical, highlighting expected returns) will select the most appropriate response at a different rate to those in Group Y (common industry terms) (Group Y ≠ Group Z).

*Hypothesis 2.3*: On average, respondents in Group Z (Technical, highlighting expected returns) will select the most appropriate response at a different rate to those in Group X (Technical, highlighting negative returns) (Group X ≠ Group Z).

### Sample selection and power calculations

#### Sample

We aimed to recruit approximately 2,000 people from the survey panel provider Dynata. Participants were eligible if they were Australian residents with a superannuation account, aged between 18 and 64 years. The sample also had quotas on age, gender and metro-regional splits so the sample was close to nationally representative along these variables. We excluded potential participants who stated that they did not currently have a super fund and did not expect to need to have a super fund in the next 12 months.

We excluded any responses that took less than 2 minutes and 30 seconds to complete the survey. This threshold was based on internal testing, prior to launching the survey, of the time to complete the survey by members of the BETA team. This was the limit below which we did not believe people could have actually been reading and engaging with the survey.

#### Power calculations

We used power calculations to determine the minimum detectable effect (MDE) for Experiment one based on an expected sample size of N = 2,000 participants*.* We set alpha at 0.1 as we were less concerned with a false positive finding. For all hypotheses in Experiment one, we conducted one-sided tests.

Based on these parameters, our calculations suggested that we would have 95% power to detect an effect size of *d* = 0.16. For the ‘choice of fund’ outcome, this equated to a difference of 0.3 scale points, assuming a pooled standard deviation of 3 scale points. For the ‘time to complete’ outcome, this represented a difference of less than a second assuming a pooled standard deviation of around 15 seconds.

### Analysis

#### Primary Analysis

For our analysis of Experiment one we used ordinary least squares (OLS) regression to estimate the effects of our intervention. The following model was used:

Where was the primary outcome variable, was a binary variable indicating if the individual was in treatment arm B or C depending on the hypotheses being assessed (the base group was arm A in both cases), was a binary covariate for education (set to 1 for university training or higher, and 0 for below university training) and is an interaction between the treatment indicator and the education covariate.

For Experiment two, we used the same model except that the treatment variable indicated whether the individual was in treatment arm X, Y or Z. Specifically:

* For hypothesis 2.1, the base group was arm X, and the treatment group was arm Y.
* For hypothesis 2.2, the base group was arm Y, and the treatment group was arm Z.
* For hypothesis 2.3, the base group was arm X, and the treatment group was arm Z.

We did not make any adjustments for multiple comparisons for either experiment however we have applied caution in how we interpret our results when multiple comparisons were conducted (see, for example, the discussion in Appendix 3 of the results for Experiment one).

#### Sub-Group and Exploratory Analysis

We also conducted sub-group and exploratory analysis. Most notably, we explored whether there was a ‘donkey vote’, that is whether participants chose the fund that appeared at the top without due consideration. We also conducted subgroup analysis on financial literacy.

### Trial threats

When writing our pre-analysis plan we were concerned some participants would fail to complete the survey and we would therefore have missing data for the primary outcomes for one or both survey experiments. However, there was only one instance of this occurring so missing data had no impact on our results.

### Limitations

#### The simplified nature of survey experiments

Experiment one was a simplified version of the comparison tool. Using the actual comparison tool to switch super fund requires multiple steps and consideration of a large amount of information and would take more time than participating in our experiment. In Experiment one, participants make a decision based on a small amount of information which could be achieved in about one minute. This simplification means that our results may have less relevance to messier real-world contexts.

Experiment one also assessed the impact of sorting products by performance category compared to a random order. In this experiment, 10 funds were shown on a single page, but in reality there many more funds that could be listed, meaning there will be a greater chance that, without sorting, the best performing funds are not visible on the first page. As such, Experiment one may underestimate the impact of sorting.

In Experiment two we presented a scenario that featured a young person. The results of this experiment may have diminished relevance in real-world situations where people are making decisions about their own superannuation.

#### Choosing an appropriate super product

In Experiment one, our outcome measure assumed that one of the ‘top’ products would always be a more appropriate product. In reality, choosing a super fund is complicated and any comparison tool will necessarily simplify some variables (such as fees or returns) or omit others (such as insurance). It is important to be mindful that the ‘top’ products will usually – but not always – represent an appropriate product, depending on an individual’s circumstances.

#### Survey sample representativeness

The extent to which the sample differs from the population of interest (that is, people with a superannuation account as well as those about to get their first account) is difficult to quantify. Instead our sample roughly matched the Australian population on several broad demographic measures. However, we drew our sample from an online survey panel. Those who sign up for an online survey panel are likely to differ from the broader population in various ways that may matter for how they think about choice of super product.

#### Survey data quality

Survey panel participants receive small incentives (for example, money or vouchers) for survey completion. Some respondents may attempt to maximise these rewards they receive by speeding through surveys without engaging with the survey content. While we attempted to reduce these issues by removing participants who completed the survey in less than two minutes, 30 seconds, the data set may still contain other disengaged respondents who gave poor quality responses.

Running two different experiments in one survey can cause issues because the treatment received in the first experiment may impact behaviour in the second experiment. Examination of treatment effects in the second experiment by first experiment treatment assignment suggests that any such impacts were negligible.

Responses to questions later in the survey may be influenced by earlier questions or survey experiments. In particular, responses to questions about performance labels preference (e.g. ‘top’, ‘strong’ or ‘high’) may have been influenced by the first experiment (at least for respondents who saw 3 performance categories, including one labelled ‘top’).

## Appendix 2: Other survey results

### Underperformance notification format preference

Respondents were asked how they would like to be notified if their fund was poorly performing. Overall, respondents displayed a clear preference for receiving the underperformance notice by email (54%) relative to mail (13%) or a combination of mail and email (33%) (Figure 10). For low financial literacy respondents, the differences were less marked: 44% preferred email relative to 22% for mail. The proportion who preferred both formats was about the same regardless of financial literacy.

### Important features for a comparison tool

Respondents stated that the most important features for a superannuation comparison site are that it be free (56% rated this as ‘very important’) and easy to use (52%) (Figure 10).

### How would you like to be reminded to check your super?

Respondents indicated that the best ways to prompt them to check the performance of their super fund would be: a letter from their super fund (39%) or a message from the Australian Government via myGov (33%) or email (31%) (Figure 11). 29% indicated that they were open to be prompted to check their super when they complete their annual tax return. Many were open to receiving prompts through SMS messages from their super fund (28%) or the Australian Tax Office (ATO) (20%). Relatively few (17%) said that prompts via TV, radio or news sites would be one of the best ways to remind them.

1. Underperformance notification format preference

Q. You are provided with a notice telling you that your superannuation product is performing poorly and directing you to the superannuation comparison tool. Which delivery method would be most likely to get your attention? n=2,236.

1. Percentage of respondents stating that a particular features was very important or quite important for a comparison tool

Q. How important to you are the following features of a comparison site, when you're comparing super funds? n=2,236. Likert scale where 1 = not important at all, 2 = not important, 3 = somewhat important, 4 = quite important, 5 = very important.

1. Prompts to check super performance

Q. What would be the best ways to remind you to check the performance of your superannuation? Select up to 3 options. n=2,236.

### Awareness campaign slogans

Respondents were asked to choose a statement that would motivate them the most to compare the performance of their super fund. While all of the suggested statements received roughly similar levels of support, the most popular (28%) stated: ‘A small change now can make a big difference in retirement’ (Figure 12). Around 20‑24% preferred one of the other three statements.

1. Motivating statements to compare your super?

Q. Which of the following statements do you think would motivate you the most to compare your super? n=2,236.

### Poor performing MySuper products

Respondents were also asked to rate the percentage of MySuper products they thought would be classed in the worst performing category. On average, respondents guessed that around 48% of products would be in the worst performing category. This is much higher than the 16% of MySuper funds that failed the Your Future, Your Super performance test in 2021.

## Appendix 3: Statistical tables for survey experiments

### Experiment one – statistical tables

1. Primary outcome – choice of fund ranking (H1.1)

| Treatment arm | Mean (ranking 1-10) | Effect (ranking 1-10) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 6.2609 |  |  |  |  |  |
| B: 3 categories, unsorted | 6.7922 | 0.5313 | 0.1474 | 0.0002 | 0.2422 | 0.8203 |

Note: n = 1,498, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – choice of fund ranking (H1.2)

| Treatment arm | Mean (ranking 1-10) | Effect (ranking 1-10) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| B: 3 categories, unsorted | 6.7898 |  |  |  |  |  |
| C: 3 categories, sorted | 7.0405 | 0.2507 | 0.1498 | 0.0472 | -0.0431 | 0.5445 |

Note: n = 1,484, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – choice of fund ranking (H1.3)

| Treatment arm | Mean (ranking 1-10) | Effect (ranking 1-10) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 6.2529 |  |  |  |  |  |
| C: 3 categories, sorted | 7.0402 | 0.7872 | 0.1490 | 0.0000 | 0.4951 | 1.0795 |

Note: n = 1,490, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – time to make a decision (H1.4)

| Treatment arm | Mean (seconds) | Effect (seconds) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 52.8492 |  |  |  |  |  |
| B: 3 categories, unsorted | 50.4410 | -2.4082 | 3.5090 | 0.2463 | -9.2913 | 4.4749 |

Note: n = 1,488, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – time to make a decision (H1.5)

| Treatment arm | Mean (seconds) | Effect (seconds) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| B: 3 categories, unsorted | 50.6148 |  |  |  |  |  |
| C: 3 categories, sorted | 51.8821 | 1.2674 | 3.6644 | 0.3647 | -5.9206 | 8.4553 |

Note: n = 1,478, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – time to make a decision (H1.6)

| Treatment arm | Mean (seconds) | Effect (seconds) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 52.7270 |  |  |  |  |  |
| C: 3 categories, sorted | 51.8851 | -0.8418 | 3.5967 | 0.4075 | -7.8970 | 6.2134 |

Note: n = 1,480, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – choice of a top fund

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 33.51 |  |  |  |  |  |
| B: 3 categories, unsorted | 55.50 | 21.98 | 2.51 | 0.0000 | 17.06 | 26.90 |

Note: n = 1,498, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – choice of a top fund

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| B: 3 categories, unsorted | 55.47 |  |  |  |  |  |
| C: 3 categories, sorted | 60.87 | 5.40 | 2.56 | 0.0174 | 0.39 | 10.41 |

Note: n = 1,484, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – choice of a top fund

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 33.48 |  |  |  |  |  |
| C: 3 categories, sorted | 60.86 | 27.38 | 2.49 | 0.0000 | 22.50 | 32.26 |

Note: n = 1,490, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – choice of any fund that is not ‘poor’ performing

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 79.92 |  |  |  |  |  |
| B: 3 categories, unsorted | 82.44 | 2.52 | 2.02 | 0.1064 | -1.44 | 6.57 |

Note: n = 1,498, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – choice of any fund that is not ‘poor’ performing

| Treatment arm | Meanv (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| B: 3 categories, unsorted | 82.42 |  |  |  |  |  |
| C: 3 categories, sorted | 83.76 | 1.34 | 1.94 | 0.2447 | -2.46 | 5.15 |

Note: n = 1,484, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – choice of any fund that is not ‘poor’ performing

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 79.81 |  |  |  |  |  |
| C: 3 categories, sorted | 83.75 | 3.94 | 1.99 | 0.0239 | 0.04 | 7.84 |

Note: n = 1,490, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – intention to switch

| Treatment arm | Mean  (5 point Likert scale) | Effect  (5 point Likert scale) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 3.5742 |  |  |  |  |  |
| B: 3 categories, unsorted | 3.5940 | 0.0198 | 0.0534 | 0.3556 | -0.0850 | 0.1246 |

Note: n = 1,498, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons. Units – 5 point Likert scale where 1 = very unlikely, 5 = very likely.

1. Secondary outcome – intention to switch

| Treatment arm | Mean  (5 point Likert scale) | Effect  (5 point Likert scale) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| B: 3 categories, unsorted | 3.6004 |  |  |  |  |  |
| C: 3 categories, sorted | 3.5926 | -0.0079 | 0.0543 | 0.5578 | -0.1143 | 0.0985 |

Note: n = 1,484, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons. Units – 5 point Likert scale where 1 = very unlikely, 5 = very likely.

1. Secondary outcome – intention to switch

| Treatment arm | Mean  (5 point Likert scale) | Effect  (5 point Likert scale) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 3.5826 |  |  |  |  |  |
| C: 3 categories, sorted | 3.5927 | 0.0101 | 0.0532 | 0.4244 | -0.0943 | 0.1146 |

Note: n = 1,490, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons. Units – 5 point Likert scale where 1 = very unlikely, 5 = very likely.

1. Secondary outcome – ease rating

| Treatment arm | Mean  (5 point Likert scale) | Effect  (5 point Likert scale) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 3.8576 |  |  |  |  |  |
| B: 3 categories, unsorted | 3.9934 | 0.1358 | 0.0476 | 0.0022 | 0.0423 | 0.2292 |

Note: n = 1,498, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – ease rating

| Treatment arm | Mean  (5 point Likert scale) | Effect  (5 point Likert scale) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| B: 3 categories, unsorted | 3.9956 |  |  |  |  |  |
| C: 3 categories, sorted | 3.8736 | -0.1220 | 0.0477 | 0.9947 | -0.2156 | -0.0284 |

Note: n = 1,484, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

1. Secondary outcome – ease rating

| Treatment arm | Mean  (5 point Likert scale) | Effect  (5 point Likert scale) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| A: 2 categories, unsorted | 3.8613 |  |  |  |  |  |
| C: 3 categories, sorted | 3.8736 | 0.0123 | 0.0486 | 0.3999 | -0.0831 | 0.1078 |

Note: n = 1,490, OLS regression, p-values are from one-tailed tests. We did not correct for multiple comparisons.

### Experiment one – further detail

The following 3 tables show the frequency with which each of the 10 super products were chosen for each of the 3 treatment arms. Key points:

* When only 2 performance categories were shown, a low-fee product (‘Acumen Super Growth’) was the single most popular product. However, its popularity fell once it was compared against other products that were rated as ‘top’ performers.
* When 3 performance categories were shown, the most popular product was ‘SafeInvest Super Growth’. This had the lowest fees amongst the 3 ‘top’ performers.
* Across the 3 arms, a similar proportion (10-13%) chose their current fund, ‘Stability Super Balanced’, which was pinned to the top of the table.

1. Choice of super product – Group A (2 performance categories, unsorted)

| Fund | Performance | Fees | Net return | Frequency | Percentage |
| --- | --- | --- | --- | --- | --- |
| Stability Super Balanced – your current fund | Poor | $892 | 6.5% | 79 | 11% |
| Gravitas Super Growth | Poor | $975 | 7.3% | 28 | 4% |
| Ultimate Super Balanced | Good | $625 | 7.4% | 65 | 9% |
| Maturity Super Growth | Good | $877 | 8.2% | 34 | 5% |
| Horizon Super Balanced | Poor | $491 | 7.1% | 44 | 6% |
| Enlightenment Super Growth | Good | $787 | 8.7% | 41 | 5% |
| Acumen Super Growth | Good | $501 | 8.2% | 220 | 29% |
| Astute Super Growth | Good | $840 | 8.9% | 123 | 16% |
| SafeInvest Super Growth | Good | $680 | 8.5% | 88 | 12% |
| Prosperity Super Growth | Good | $720 | 7.9% | 30 | 4% |

Note: n = 752. Fund order in this table reflects the order they were presented to participants in this experimental group.

1. Choice of super product – Group B (3 performance categories, unsorted)

| Fund | Performance | Fees | Net return | Frequency | Percentage |
| --- | --- | --- | --- | --- | --- |
| Stability Super Balanced – your current fund | Poor | $892 | 6.5% | 71 | 10% |
| Gravitas Super Growth | Poor | $975 | 7.3% | 32 | 4% |
| Ultimate Super Balanced | Fair | $625 | 7.4% | 52 | 7% |
| Maturity Super Growth | Fair | $877 | 8.2% | 44 | 6% |
| Horizon Super Balanced | Poor | $491 | 7.1% | 28 | 4% |
| Enlightenment Super Growth | Top | $787 | 8.7% | 103 | 14% |
| Acumen Super Growth | Fair | $501 | 8.2% | 84 | 11% |
| Astute Super Growth | Top | $840 | 8.9% | 132 | 18% |
| SafeInvest Super Growth | Top | $680 | 8.5% | 179 | 24% |
| Prosperity Super Growth | Fair | $720 | 7.9% | 21 | 3% |

Note: n = 746. Fund order in this table reflects the order they were presented to participants in this experimental group.

1. Choice of super product– Group C (3 performance categories, sorted)

| Fund | Performance | Fees | Net return | Frequency | Percentage |
| --- | --- | --- | --- | --- | --- |
| Stability Super Balanced – your current fund | Poor | $892 | 6.5% | 94 | 13% |
| Enlightenment Super Growth | Top | $787 | 8.7% | 91 | 12% |
| SafeInvest Super Growth | Top | $680 | 8.5% | 199 | 27% |
| Astute Super Growth | Top | $840 | 8.9% | 158 | 21% |
| Acumen Super Growth | Fair | $501 | 8.2% | 96 | 13% |
| Maturity Super Growth | Fair | $877 | 8.2% | 28 | 4% |
| Prosperity Super Growth | Fair | $720 | 7.9% | 22 | 3% |
| Ultimate Super Balanced | Fair | $625 | 7.4% | 23 | 3% |
| Horizon Super Balanced | Poor | $491 | 7.1% | 20 | 3% |
| Gravitas Super Growth | Poor | $975 | 7.3% | 7 | 1% |

Note: n = 738. Fund order in this table reflects the order they were presented to participants in this experimental group.

### Experiment two – statistical tables

We pre‑specified 3 hypotheses to test differences between the 3 pairs of arms in Experiment two (H2.1-H2.3). Group Y used common industry terms (for example, ‘conservative’, ‘balanced’ or ‘growth’ to describe the risk‑return trade‑off. As reported, we found that more people selected the optimal investment strategy in this group than in the others (where accurate but technical language was used). This meant that we rejected the null hypothesis for H2.1 and H2.2 but we were unable to do so when comparing Groups X (Technical definition, highlighting negative returns) and Z (Technical definition, highlighting expected returns (tables 22-24). The average treatment effects were large (19-22%) and would be of material significance if they held in the real world.

1. Primary outcome – selection of the most appropriate response (H2.1)

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| X: Technical, highlighting negative returns | 27.26 |  |  |  |  |  |
| Y: Common industry terms | 49.04 | 21.78 | 2.45 | 0.000 | 16.97 | 26.58 |

Note: n = 1,494 OLS regression, p-values are from two-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – selection of the most appropriate response (H2.2)

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| Y: Common industry terms | 49.04 |  |  |  |  |  |
| Z: Technical, highlighting expected returns | 30.24 | -18.81 | 2.49 | 0.000 | -23.69 | -13.92 |

Note: n = 1,484, OLS regression, p-values are from two-tailed tests. We did not correct for multiple comparisons.

1. Primary outcome – selection of the most appropriate response (H2.3)

| Treatment arm | Mean (%) | Effect (%) | Std. error | p-value | 95%CI: low | 95%CI: high |
| --- | --- | --- | --- | --- | --- | --- |
| X: Technical, highlighting negative returns | 27.35 |  |  |  |  |  |
| Z: Technical, highlighting expected returns | 30.37 | 3.02 | 2.34 | 19.77 | -1.57 | 7.61 |

Note: n = 1,494, OLS regression, p-values are from two-tailed tests. We did not correct for multiple comparisons.

### Respondent demographic characteristics

This section outlines the demographic characteristics of the overall sample. For more details on sample demographics see the spreadsheet appendix.

The survey sought to capture a sample reflective of the Australian population based on age, gender and metropolitan-regional split. Table 25 shows this was broadly achieved, though the share of respondents aged 18-24 is lower than that targeted. This meant that our sample was slightly overweight in older age groups. The lower proportion of younger respondents is likely a product of the survey being administered online in a form incompatible with mobile devices, which made younger demographics more difficult to reach. Additionally, respondents were required to have a superannuation account (or expect to have one within a year) and not yet be retired, which is also less likely at younger ages.[[5]](#footnote-6)

With respect to education, 55% of the sample had achieved a bachelor degree or higher qualification. The majority (63%) of respondents reported being in full‑time employment, with another 27% in part‑time or casual employment.

Only 15% of respondents reported having two or more superannuation products. This is significantly lower than the 26% reported by the ATO (2021) using administrative data.

1. Table 25: Representativeness of sample relative to targets based on the 2016 census

|  | Target % | Achieved % |
| --- | --- | --- |
| **Gender** |  |  |
| Male | 49% | 49% |
| Female | 51% | 51% |
| **Age** |  |  |
| 18-24 | 16% | 10% |
| 25-34 | 23% | 22% |
| 35-44 | 22% | 25% |
| 45-54 | 21% | 22% |
| 55-64 | 19% | 21% |
| **Location** |  |  |
| Metropolitan | 69% | 68% |
| Regional | 31% | 32% |

In terms of engagement with superannuation:

* 48% reported having previously switched fund
* Nine per cent reported having switched fund in the past twelve months, with another 29% stating they had considered doing so
* 74% stated they had previously researched superannuation funds, spoken to friends or family about which superannuation funds are the best, visited a superannuation comparison site or asked for superannuation advice from an accountant or financial planner.

On a set of three questions commonly used to measure financial literacy, 38% of respondents selected one or zero correct answers. This is higher than the 30% seen in the PC’s survey of superannuation fund members completed as part of their inquiry (PC 2018a:56).

## Appendix 4: Survey wording

### Overview

The complete list of survey questions, including the 2 survey experiments are below. They are presented in the order that respondents saw them.

### Survey questions

#### Demographics – part 1

1. **Please select your age bracket**

* 18 - 24
* 25 - 34
* 35 - 44
* 45 - 54
* 55 - 64
* 65+

1. **What is your gender?**

* Female
* Male
* Non-binary
* Prefer not to say

1. **What is the postcode where you usually live?**
2. **What is the highest level of education that you have completed?**

* Year 10 or below
* Year 11 or equivalent
* Year 12 or equivalent
* A trade, technical certificate or diploma
* A university degree
* Postgraduate qualifications

1. **Which of the following best describes your current employment status?**

* Working full time
* Working part time
* Working casually
* Self-employed / Business owner
* Not currently working / unemployed
* Student
* Retired
* Home duties including caring for others
* Unable to work due to illness, disability or impairment
* Other (please specify)

#### Survey Experiment one

This study is about comparing superannuation options. We’re going to ask you to respond to an imaginary situation in which you’re comparing super funds. Try to answer the questions as you would in a real experience like this.

Imagine you’ve been working for about a decade- you’re 32 years old. You’re reading a news article on your laptop that mentions a new Government website - a superannuation comparison tool. The article also says that you can change your super fund if you want to. You follow a link, and go to the site.

It prompts you to enter a few details, which you do. You put in your age (32), your current super fund (Stability Super), and your best estimate of how much super you currently have ($75,000).

Click through to the next page to see your personalised results.

*[Respondents saw one of the three results tables, as shown in the ‘What we did’ section]*

1. **Which one do you think looks like the best option?**

* Stability Super Balanced - your current fund
* Gravitas Super Growth
* Ultimate Super Balanced
* Maturity Super Growth
* Horizon Super Balanced
* Enlightenment Super Growth
* Acumen Super Growth
* Astute Super Growth
* SafeInvest Super Balanced
* Prosperity Super Growth

1. **When deciding which fund looked best, which feature did you focus on the most?**

* Easiest option
* Lowest fees
* Best performance
* Lowest risk
* Highest returns
* Liked the name
* Unsure
* Other (please tell us)

1. **Remembering that you're answering for your 32 year old self, how likely is it that you would switch funds?**

* Very likely
* Likely
* Neither likely nor unlikely
* Unlikely
* Very unlikely

1. **How easy was it to compare super funds using the superannuation comparison tool?**

* Very easy
* Easy
* Neither easy nor difficult
* Difficult
* Very difficult

#### Survey Experiment two

1. **Please consider the following descriptions of super fund performance. Which label would be most helpful in highlighting the best performing funds?**

* Top
* Strong
* High

1. **And which label would be most helpful in highlighting the worst performing super funds?**

* Under
* Poor
* Weak

1. **There are 88 different default superannuation funds available in Australia. What percentage of superannuation funds do you think would be worst performing? Please give us your best guess.**
2. **In the process of signing up with a new super fund, you are offered a choice of investment options. Which investment option are you most likely to choose?**

[*Respondents saw one of the following three sets of response options.]*

* Low risk – expect a negative return 1 out of every 20 years
* Medium risk – expect a negative return 2 out of every 20 years
* High risk – expect a negative return 4 out of every 20 years

*or*

* Conservative
* Balanced
* Growth

*or*

* Low risk & low expected return
* Medium risk & medium expected return
* High risk & high expected return

1. **Which is the best option for a young person starting out in their career?**

*[Respondents saw one of the following three sets of response options.]*

* Low risk – expect a negative return 1 out of every 20 years
* Medium risk – expect a negative return 2 out of every 20 years
* High risk – expect a negative return 4 out of every 20 years

*or*

* Conservative
* Balanced
* Growth

*or*

* Low risk & low expected return
* Medium risk & medium expected return
* High risk & high expected return

#### Survey questions

1. **You are provided with a notice telling you that your superannuation product is performing poorly and directing you to the superannuation comparison tool. Which delivery method would be most likely to get your attention?**

* By mail
* By email
* By both mail and email

1. **Which of the following statements do you think would motivate you the most to compare your super?**

* Make your super work harder for you
* Make the most of your super
* Take control of your super
* A small change now can mean a big difference in retirement
* None of the above
* Other

**Why?**

1. **When do you think you would be most likely to use a superannuation comparison tool? Select all that apply.**

* When I start a new job or change jobs
* When I get my annual statement from my super fund
* If I got a letter or email from my super fund
* If my accountant or financial adviser suggested it
* If I heard about it from friends/family
* If I heard about it on social media
* If I heard about it on TV, radio, newspaper or a news website
* When I do my tax return
* When I get a pay increase
* When I'm close to retiring
* Other (please explain)

1. **How important to you are the following features of a comparison site, when you're comparing super funds?** (*Response options:**Not important at all, Not very important, Somewhat important, Quite important, Very important*)

* Easy to use
* Free - no cost to use
* The information is from an independent source (i.e., Government)
* It can be personalised with my current super information if I login

1. **What would be the best ways to remind you to check the performance of your superannuation? Select up to 3 options.**

* An SMS message from the Australian Taxation Office
* An SMS message from my super fund
* A message in myGov
* A letter from the Australian Government
* A letter from my super fund
* An email from the Australian Government
* TV, radio, newspapers, or news sites
* A reminder when I'm doing my tax return
* I don't want to be reminded
* Other (please explain)

#### Demographics – part 2

1. **What is your personal annual income from all sources before tax? Please include all wages, salaries, pensions and other income. If you are unsure, your best guess will be fine.**
2. **Which of the following statements comes closest to describing the amount of financial risk that you are willing to take with your spare cash? That is, cash used for savings or investment.**

* I take substantial financial risks expecting to earn substantial returns
* I take above-average financial risks expecting to earn above-average returns
* I take average financial risks expecting to earn average returns
* I take below-average financial risks expecting to earn below- average returns
* I am not willing to take any financial risks
* I rarely have any spare cash

1. **What is your current superannuation balance (approximately)? If you have multiple funds please combine the balances and report the total amount.**

* Less than $10,000
* $10,000 - $49,999
* $50,000 - $99,999
* $100,000 - $199,999
* $200,000 - $499,999
* $500,000 +
* Don't know

1. **How many super funds do you have?**

* 0
* 1
* 2 or more
* Don't know

1. **Which of the following best describes your attitude towards superannuation?**

* I don't know a great deal about financial products other than transaction accounts and credit cards
* I don't take much interest in my superannuation, I am not even sure how much I have invested/what sort of investment mix I have
* I know a bit about my superannuation fund but don't think about superannuation much in between getting my annual statements
* I try to keep up with how my superannuation is going and check from time to time whether I have the right sort of investment options
* Superannuation is an important part of planning for my future, I regularly review it to get the best performance I can

1. **Have you ever switched your superannuation from one fund to another?**

* Yes
* No
* Don't know / can't remember

1. **In the last 12 months, have you considered switching your super balance to another fund?**
   * Yes - have considered
   * Yes - already switched
   * No
   * Not sure
2. **Have you ever done any of the following? Please select all that apply.**

* Visited a superannuation fund comparison site
* Asked for advice from an accountant or financial planner about to your superannuation
* Done your own research on superannuation funds
* Talked to friends or family about which superannuation funds are best
* Used a superannuation/retirement savings calculator
* None of the above

1. **Which of the following best describes your approach to investing for retirement?**

* Use my employer's default fund
* Choose a low risk option
* Consider the number of years to retirement
* Choose an ethical or sustainable option
* Choose the option that will grow the fastest
* Balance the risk and returns
* Don't know
* Other
* Does not apply to me

1. **Suppose you put $100 into a no-fee savings account with a guaranteed interest rate of 2% per year. You don't make any further payments into this account and you don't withdraw any money. How much would be in the account at the end of the first year, once the interest payment was made?**

* More than $102
* Exactly $102
* Less than $102

1. **Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After one year, would you be able to buy more than today, exactly the same as today, or less than today?**

* More than today
* Exactly the same as today
* Less than today

1. **Do you think the following statement is True or False?***"Buying shares in a single company usually provides a safer return than buying shares in a number of different companies."*

* True
* False

1. **If you have any comments about the survey or this topic, please provide them here.**

## References

Australian Government the Treasury (2020), *Retirement Income Review: Final report,* accessed 1 August 2022. <https://treasury.gov.au/sites/default/files/2021-02/p2020-100554-udcomplete-report.pdf>

Australian Prudential Regulation Authority (2021), *Annual MySuper statistics 2021*, accessed 12 July 2022. <https://www.apra.gov.au/annual-mysuper-statistics>

Australian Prudential Regulation Authority (2022), *APRA releases superannuation statistics for March 2022*, accessed 12 July 2022. https://www.apra.gov.au/news-and-publications/apra-releases-superannuation-statistics-for-march-2022

Australian Taxation Office (2021), *Super data: multiple accounts, lost and unclaimed super*, accessed 12 August 2022. : https://www.ato.gov.au/About-ATO/Research-and-statistics/In-detail/Super-statistics/Super-accounts-data/Multiple-super-accounts-data/

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman (2015). "Choice overload: A conceptual review and meta-analysis." Journal of Consumer Psychology 25, no. 2: 333-358.

Lusardi, Annamaria, and Olivia S. Mitchell (2014). "The economic importance of financial literacy: Theory and evidence." Journal of economic literature 52, no. 1 (2014): 5-44.

Productivity Commission 2018a, *Superannuation:* *Assessing Efficiency and Competitiveness*, Report no. 91, Canberra.

Productivity Commission 2018b, ‘Members surveys’, Technical Supplement 1 to the Inquiry Report Superannuation: Assessing Efficiency and Competitiveness, Canberra, December.

**© Commonwealth of Australia 2022**

ISBN 978-1-925365-14-6 YourSuper Comparison Tool: Results from a survey and two survey experiments (online)

**Copyright Notice**

With the exception of the Commonwealth Coat of Arms, this work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)



**Third party copyright**

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material. Please contact them directly.

**Attribution**

This publication should be attributed as follows:

© Commonwealth of Australia, Department of the Prime Minister and Cabinet, *YourSuper Comparison Tool: Results from a survey and two survey experiments*.

**Use of the Coat of Arms**

The terms under which the Coat of Arms can be used are detailed on the following website: <https://pmc.gov.au/cca>



Behavioural Economics Team   
of the Australian Government

General enquiries [beta@pmc.gov.au](mailto:beta@pmc.gov.au)

Media enquiries [media@pmc.gov.au](mailto:media@pmc.gov.au)

Find out more [www.pmc.gov.au/beta](http://www.pmc.gov.au/beta)

1. In Group A (2 categories, unsorted) 20% selected the ‘poor’ category, compared to 17-18% for Group B (3 categories, unsorted) and Group C (3 categories, sorted). 46% of ‘Group A chose a ‘middle-ranking’ product, which fell to 27% and 23% respectively for Group B and Group C. See tables 10‑12 in Appendix 3 for more details. [↑](#footnote-ref-2)
2. For Group A (2 categories, unsorted) ‘Acumen Super’ had the lowest fees amongst the ‘good’ products but ranked fourth on performance. For Groups B and C (both with 3 categories), ‘SafeInvest Super’ was the lowest fee option of the ‘top’ products but it ranked second on performance. [↑](#footnote-ref-3)
3. We constructed a measure of respondent financial literacy using responses to 3 questions on knowledge of interest rates, inflation and investment diversification. The questions were adapted from the ‘Big 3’ set devised by Lusardi and Mitchell (2014). Respondents who scored one or below were deemed to have low financial literacy (38% of the sample). [↑](#footnote-ref-4)
4. The spreadsheet appendix is available at: <https://behaviouraleconomics.pmc.gov.au/projects/yoursuper-comparison-tool-results-survey-and-two-survey-experiments> [↑](#footnote-ref-5)
5. One respondent who had retired was included in the sample owing to a decision to screen on this criterion being made after the soft launch of the survey occurred. [↑](#footnote-ref-6)