Submission to YFYS Consultation

April 2024

About Chant West

Chant West is a leading superannuation research firm established in 1997. It conducts research on most of the leading superannuation and pension funds in Australia. Its research is purchased by most of Australia's leading superannuation suppliers and its comparison tools are widely used by consumers, funds and financial advisers. Chant West was purchased in June 2020 by Zenith Investment Partners and is now part of the FE fundinfo group.

Introduction

This submission provides Chant West's feedback on Treasury's 'Annual Superannuation Performance Test – design options' consultation paper that was issued in March 2024.

We support the use of a performance test to assess the quality of outcomes delivered by superannuation funds to their members, and to identify underperforming funds, particularly for MySuper products where members do not make an active decision to invest in the product. We also acknowledge that it is notoriously difficult to construct a test that measures performance in a way that accounts for the different levels of risk in each portfolio. In this submission, we provide feedback on Treasury's proposed design options and indicate a preference for a multi-metric test that complements the current performance test to ensure that underperforming funds are more reliably identified.

Problems with the current performance test

The current performance test measures how well a fund has implemented its strategic asset allocation by comparing to a reference portfolio with the same asset allocation that returns benchmark performance for each asset class. This test measures the effectiveness of implementation, but it does not have a strong correlation to member outcomes, which is mainly determined by the strategic asset allocation itself.

The combination of the low efficacy of the test in identifying true underperformance, and the dire consequences of failure, is highly problematic and has led to significant changes in the way many funds invest. The existential threat of failing the test has led some funds, especially those close to failing the test, to focus more on passing the test in the current year rather than focussing on long-term returns (i.e. a shift to a short-term focus). This is not in the interests of super fund members. Indeed, some funds terminated strategies in 2021 with relatively high YFYS tracking error that would have provided very effective protection in the FY22 investment environment. And other funds have turned down very good investment opportunities that didn't align well with the YFYS benchmarks. The result has been poorer returns for members and inferior member outcomes in these funds.

We understand the perspective of many funds that are happy for the current test to remain – 'better the devil you know' – as since 2021 they have been actively managing to the test. Indeed, the only MySuper products to fail the test in 2022 and 2023 were those that failed in 2021 (or another division of those products). Many super funds are happy to retain the test as they know they are almost assured to not fail as they actively manage to the test.

Our concern is that some funds focus too much on passing the test rather than providing the best investment return for the level of risk taken, which is what they should be focussed on.

Comments on Consultation questions 2-8

- The current test could be marginally improved by adding more benchmarks to better align with the way that some funds invest, including catering for the energy transition and other emerging asset sectors. However, since each fund has different objectives and strategies for each asset class, the number of benchmarks required to do this effectively would be very large indeed.
- Adding more indices will mean higher costs for superannuation funds and service providers with index providers. The prohibitive fees currently charged by index providers for the mandated indices ultimately detract from member outcomes across the industry, and will further do so if more indices are added. We believe it is difficult for index providers to justify the level of fees currently charged for the required indices. A reversion to more standard listed indices may help reduce this cost with possibly little change to performance test outcomes.
- The suggestion to move from a large number of benchmarks to a small number of listed benchmarks in a Simple Reference Portfolio is problematic as there will be some periods when these assets outperform more diversified portfolios and many funds may fail the test over that period as a result.

The best way forward?

Section 2 of the consultation paper outlines a few alternative metrics that could be used as alternatives to the current test. However, each of these tests have flaws of their own that are accurately detailed in the ‘Benefits and drawbacks’ included for each metric. This means that none of these tests are ideal for a single bright line test. Indeed, replacing the current performance test with an alternative bright line test would introduce another round of retrospectivity, which was a particularly unfortunate feature of the original iteration of the performance test in 2021.

This consideration pushes us down the path of a multi-metric approach. Indeed, multiple metrics would provide different perspectives on how a fund has performed and would provide much greater efficacy in recognising underperformance, in contrast to the current test’s myopic assessment. Multiple metrics would provide a much greater level of certainty on whether a fund is truly an underperforming fund.

Treasury’s potential candidates for a multi-metric test are outlined in Sections 2 and 3 of the consultation paper. Before we discuss the options that we believe are viable for inclusion, we will eliminate those metrics we believe would be unsuitable for inclusion, as follows:

- **Sharpe Ratio –** the Sharpe ratio metric would be difficult to explain to members, it does not come with any natural benchmark to define a pass/fail assessment and any pass/fail ratio would need to be different for different levels of growth assets due to way it is calculated.
- **Administration fee –** we do not believe that one of the limbs of the multi-metric test should be an assessment of administration fees as the YFYS test is explicitly a performance test. Indeed, if such a test was included that required that RAFE be say less than the benchmark RAFE less 50 bps, there would be an incentive for funds to increase administration fees to just below that amount and to reduce investment fees to maximise the probability of passing the performance component of the test. We believe that administration fees should be deducted from the investment performance that is used for the performance test and not assessed separately.

- **CPI+X** – the CPI+X metric is problematic as there will be times when all portfolios will pass the test and other times when all portfolios will fail the test. Chart 1 compares the rolling 10 year median performance of growth funds (61-80% growth assets) with a typical CPI+3.5% investment objective. It shows that all growth funds would have passed the test for most of the period from the introduction of compulsory superannuation as the median return is well above the objective. But it also shows that most funds would have failed the test for the four years from 2008 to 2012 and from 2015 to 2017 (as the median is below the objective for these periods). This illustrates this metric's limitations in identifying underperforming funds, as there will be periods when all products pass and others when most products fail – this makes it highly inappropriate to include in any test.

Chart 1 - Growth Funds (61-80% growth assets) – Rolling 10 Year Performance (Returns % pa)

We believe that the following two metrics would be appropriate candidates for a supplementary test as they measure real member outcomes by assessing the return provided for a given level of risk (or proxy of risk, i.e. growth assets). This is aligned with the objective of any investor which is to provide a strong return for the level of risk taken.

- **Peer comparison – 2(b)** – the performance of each portfolio could be compared with other portfolios (either all MySuper or all trustee-directed products) using a line of best fit across different levels of growth assets to identify funds with poor returns (say 50 bps below the line of best fit). The data for this test would be the return over the period (probably net investment return less current administration fees) as well as the average growth assets over the period (using the growth assets from APRA's quarterly data). While this metric is well-aligned to member outcomes, it does have some limitations:
    - This approach relies on determining the level of growth assets for each portfolio. While APRA has proposed a simple method for determining growth assets that it uses in its heatmap, there are problems with it (especially for alternatives and unlisted assets). This peer comparison approach would be greatly assisted if there was a widely accepted and widely used definition of growth assets, which is proposed or supported by APRA.

  - A peer comparison does not provide insight into whether all funds are performing well, or whether all funds are performing poorly. Indeed, it is likely that some funds will always fail such a test even though all funds may be performing well compared with other multi-asset portfolios not in MySuper or indeed outside superannuation.
  - As the next section will show, far more conservative portfolios fail this metric, which suggests that a 'line' of best fit may not be most appropriate to represent the 'median' fund. However, a shift to use a 'curve' of best fit would further complicate this approach.

- **Risk-adjusted returns relative to simple reference portfolio – 2(c)** – the performance of each MySuper or trustee-directed product could be compared with the risk-adjusted return of a Simple Reference Portfolio with the same level of risk. The Simple Reference Portfolio represents a naïve passive portfolio comprised simply of listed equities, bonds and cash (as defined in the Productivity Commission's final report. This approach measures the value added by the fund over and above such a 'no skill' portfolio – a fund should be able to at least match that return. The data for this test is readily accessible as it is simply the quarterly MySuper return data, from which the performance and volatility over the period can be calculated – this could be adjusted by reducing the return by the current administration fees (RAFE) for the product and the benchmark return by the benchmark RAFE. While this metric is also well-aligned to member outcomes, it also has some limitations:
  - This approach does favour those funds with larger exposures to unlisted assets which will have lower volatility due to the frequency of valuations (these assets are generally valued quarterly, rather than daily for listed markets). In this way, funds would be incentivised to have more diversified portfolios including a meaningful amount of unlisted assets, which may not be a bad outcome, but this would be difficult for advised trustee-directed products which have a very different liquidity profile due to much greater engagement by advisers.
  - While some may be concerned that this approach will lead to funds valuing unlisted assets less often, we do not believe this will be the case as APRA has been driving funds to more frequent valuations and will continue to do so. We have observed clear evidence of enhanced valuation practices over the last couple of years..
  - Also, while volatility is a useful metric for risk that can be easily calculated at any point in time, it only reflects one type of risk.
  - The results of this test will partly depend on the precise composition of the Simple Reference Portfolio, including such items as whether REITs are included (and whether they are Australian and/or global), the mix of Australian and international assets in each asset class and the mix between fixed interest and cash.

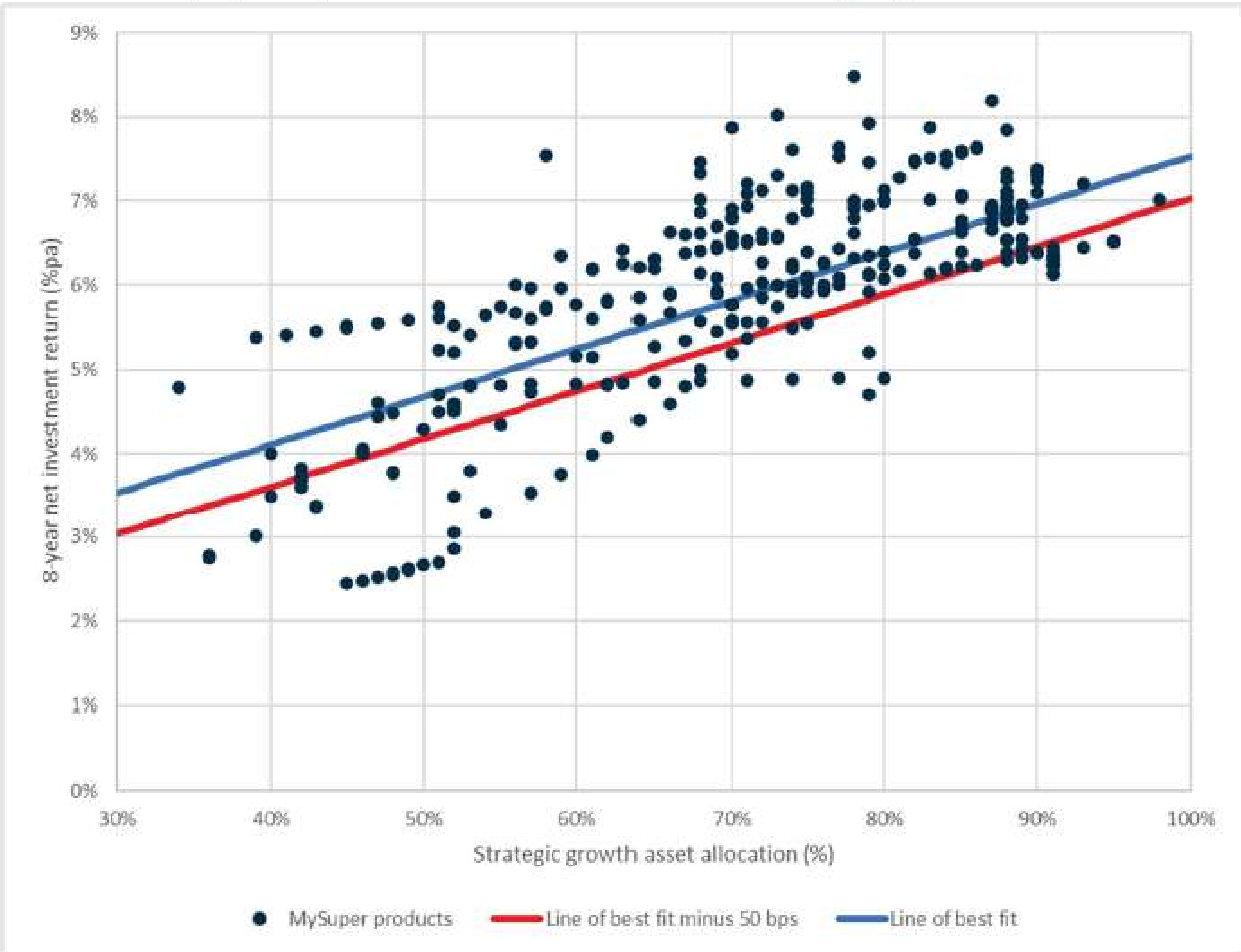
Results of these additional tests?

In order to assess how well these metrics would work as part of an enhanced multi-metric performance test, we need to understand the results of applying these two supplementary metrics to existing funds. In this section, we do that using data from the consultation paper which is based on APRA data.

A. Peer comparison – 2(b)
Chart 2 in the consultation paper, which is replicated below, shows the net investment return versus strategic growth allocation for the 8 years to 30 June 2022, with the blue line showing the line of best fit and the red line showing the line of best fit less 50 bps. The MySuper products represented by the blue dots below the red line that would have failed this test metric.

When we look closely at the data behind each 'blue dot', while many blue dots appear below the red line, they mainly represent some older age cohorts of a small number of lifecycle products. Significantly, most of the younger age cohorts in these products are above the red line and, given there are far more younger members than older members, most of these funds would pass this metric overall. Our calculations suggest that only two MySuper products would have failed this test metric – a retail product that failed the current performance test and another small retail lifecycle product that passed the current performance test. The industry funds that failed the current performance test would have passed this test. So, in summary, fewer funds would have failed this test than with the current performance test.

Chart 2 – MySuper 8-year net investment return vs strategic growth asset allocation

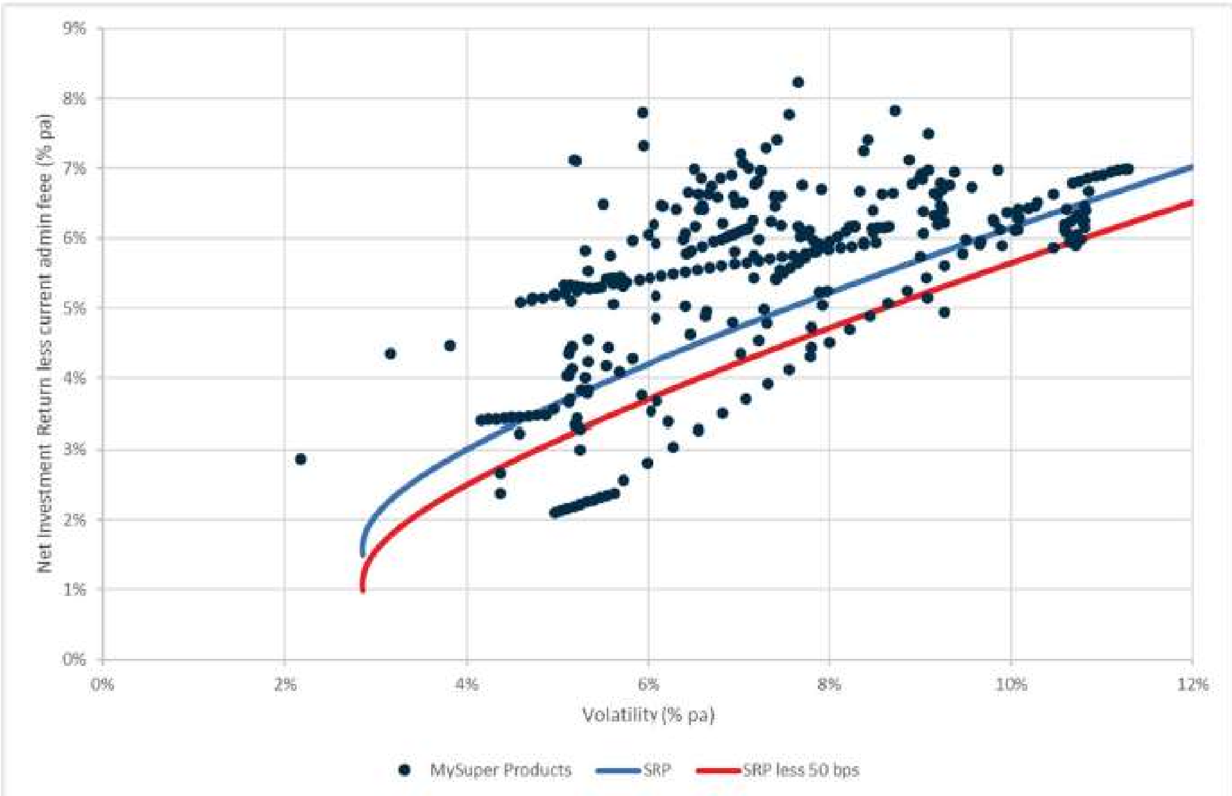


Note: a linear line of best fit is used. Analysis also shows multiple results for lifecycle products as each life stage is represented separately.
Source: APRA 2022 MySuper Heatmap.

B. Risk-adjusted returns relative to simple reference portfolio – 2(c)
Chart 3 in the consultation paper, which is replicated below, shows the net investment return less current administration fees versus volatility based on APRA's MySuper performance quarterly data for the 8 years to 30 June 2022. The blue line shows the return vs volatility for all possible Simple Reference Portfolios and the red line showing the SRP return less 50 bps. Once again, it would be the MySuper products represented by the blue dots below the relevant line that would have failed this test metric.

Under this test, if no tolerance was provided below the benchmark return (i.e. the blue line applies), all products except one that failed the current performance test for the 8 years to 30 June 2022 would also have failed this test. In addition, four retail products made up almost solely of listed assets would also have failed. If a 50 bps tolerance was included (i.e. the red line applies), only 1 product would have failed the test. As was stated earlier, this test does benefit those funds with significant levels of unlisted assets as the valuation frequency of these assets reduces the overall volatility and thereby reduces the return target. To adjust for this, there may be an argument to apply different thresholds based on the level of unlisted assets. For example, for products that are almost all listed (this will include lots of trustee-directed products offered in adviser-led choice products), the 50 bps tolerance could be applied but for those funds with, say, more than 10% unlisted assets, then a smaller tolerance could be applied.

Chart 3 – MySuper Return vs Risk (8-year)



Note: Analysis also shows multiple results for lifecycle products as each life stage is represented separately.
Source: data sourced from Chant West

How could these tests be combined with the current test?

The current performance test can produce results which incorrectly identify a fund as underperforming which may be providing good member outcomes but fail the test due to their investments in some asset sectors not closely matching the relevant benchmarks (perhaps due to a desire to reduce risk in these asset classes). The introduction of a supplementary test could provide another perspective on fund performance to confirm whether a fund that fails the current test is truly an underperforming fund.

Two test structure

If two tests were applied, we believe the following would be most appropriate:
- Current performance test metric
- Performance relative to the relevant Simple Reference Portfolio – 2(c)

This would mean that if a MySuper (or choice) product failed both metrics, it would be deemed to have failed the test and would bear the consequences. However, if it failed just one of the tests then it would be deemed not to have failed. While it may appear that fewer funds would fail such a performance test, we expect few funds will fail the test in coming years anyway, as funds have shown they are quite good at managing to the current test (now that they know what it is). In addition, we believe that the first two years of the test have already been effective in removing the chronic underperformers (although it probably removed a couple of others as well).

Even though the current metric would be retained as part of the performance test, we believe that the addition of the alternative metric will mean that funds will not have to be so focussed on passing the current metric (with its existential consequences) but rather focus more on providing strong risk-adjusted returns, which is what is in members' best interests. This would free up funds to invest for long-term risk-adjusted returns and incorporate their best ideas into portfolios, even when they introduce greater tracking error. This would include the freedom to invest in the energy transition and emerging asset sectors for which there are not relevant benchmarks.

We do not believe that the goal should be to produce more failures of the test, but rather to ensure that when a MySuper is identified as an underperforming fund, that it is indeed an underperforming fund. We don't believe that the addition of the supplementary metric weakens the test but rather that it makes the result of the test more reliable and less susceptible to 'false positives', which was a problem with the current test.

Three test structure

However, we understand that allowing a fund to pass one of two tests may be perceived (and possibly portrayed) as a weakening of the test – and this may be politically untenable. If that was the case, it may be necessary to include two supplementary tests so that a total of three tests are applied, and a fund would need to pass two out of three tests to pass. We would not recommend that the first test is the primary test but that each test has equal priority – to reduce any anchoring to current behaviour which ensures that the current test is passed at all costs.

If a three test structure is used, we believe that the tests should comprise the following:
- Current performance test metric
- Performance relative to the relevant Simple Reference Portfolio – 2(c)
- Peer relative test – 2(b)

One downside of this three test structure is that it introduces some retrospectivity, as a fund may have built its investment strategy around ensuring that it can pass the current test, but it may actually fail the other two tests and therefore fail the performance test overall.

Technical working group

The preceding discussion includes a number of points where more work may be required to clarify details of any new approach including appropriate thresholds to use for each test, the composition of Simple Reference Portfolio, additions/changes to benchmarks used in the current metric etc. We would recommend that this working group be comprised of members with as little conflicts as possible to ensure that outcomes are determined based on objective analysis.

Administration fees and costs – period used & platform products

We support the inclusion of administration fees and costs in the current performance test and any other limbs of a multi-metric test. We believe that the use of current/recent administration fees and costs is appropriate as it is a much better indication of ongoing member outcomes in these products. More importantly, this approach has directly led to significant reductions in administration fees in some funds that have benefitted members. We would be open to retaining the current 'last 12 months' fees or extending to the last 2-3 years, which would limit the ability to impact the performance test result in a particular year simply by slashing administration fees.

Further, we support a separate RAFE for platform products as they provide a very different service offering to other superannuation products, including a wide-ranging investment menu, sophisticated features and a high quality adviser portal to help advisers efficiently service their clients. However, the use of a $50,000 balance for platform products is inappropriate and leads to misleading results. We collect annual data on the major platforms and at 30 June 2023, there was a total of about $230 billion assets in wrap products (super & pension) for about 900,000 members which gives an average balance of just over $250,000. For this reason, it would be much better to base each platform's RAFE (and the benchmark RAFE) on a $250,000 balance. Indeed, several platform products have a higher $-based administration fee and lower %-based fee to better reflect the actual costs of administration, but these products have been greatly penalised by the use of a $50,000 balance for the RAFE calculation (such products are not intended for members with such balances). **A balance of $250,000 should be used for the platform RAFE calculation.**

Choice options

The range of choice options is much wider than the range of MySuper products, so the problems with the current test are magnified when applied to choice options. However, we still believe that diversified choice options (at least trustee-directed options) should be included in the performance test, especially as some of these options charge higher fees which may lead to worse member outcomes. But we also need to remember that choice products have been selected by members and their advisers, unlike MySuper products which are the default options in each product. This difference suggests that the consequences of failure should probably be different. We believe that failure to pass an initial test metric should not necessarily lead to the significant consequences of the current test. The consequences could be changed to require, say, three years of failures to bear the consequences of failure or it could enter a review process with APRA that considers a range of other metrics that take into account the particular objectives of the portfolio.

We believe that single sector investment options should be excluded from the performance test. These options are typically used in a portfolio of sector options across multiple asset classes as part of an adviser's investment strategy for a client or group of clients. The only way to apply the current test to single sector options would be to use the specific benchmarks in each portfolio's PDS but this is what all funds are required to do under SPS530 anyway – so we don't believe the performance test would add anything of value for these products.

Finally, we believe that retirement products (i.e. account-based pensions and the developing longevity products) should be excluded from the performance test in the near term. These portfolios are often ill-suited to the current performance test as these portfolios often take positions within asset classes to reduce risk (as these members are in drawdown). These positions may show up as underperformance in the current test during times of strong markets (like in the first seven years of the performance test) but they are actually examples of the product provider tailoring portfolios to better suit pension members.

With the introduction of the Retirement Income Covenant, the last thing we want to do is introduce disincentives for funds to tailor solutions for pension members to help maximise retirement income which managing the risk to the stability of that income. Rather, we need to provide funds with incentives to innovate and create better solutions for pension members.