Level 15, 360 Collins Street
Melbourne VIC 3000

PO Box 18041,
Collins Street East VIC 8003

T 1300 300 820 Member Services
1300 017 589 Retirement Services
1300 304 947 Employer Services

E vision@visionsuper.com.au
W www.visionsuper.com.au

Superannuation, Efficiency and Performance Unit
Retirement, Advice and Investment Division
The Treasury

By email: YFYS@treasury.gov.au

19 April 2024

Dear Superannuation team

Annual superannuation performance test – design options consultation paper

Introduction

Vision Super is a mid-sized superannuation fund, with 85,000 member accounts and around $13 billion in funds under management. Vision Super was founded in 1947, to look after retirement benefits for workers in the local government and authorities sectors in Victoria. We retain a strong and significant connection with our members in our traditional sectors. We have looked after members' retirement savings and pensions for over 75 years, managing the full range of benefit designs from MySuper and choice accumulation products, and closed defined benefit schemes to lifetime and allocated pensions.

Current Your Future Your Super performance test

From Vision Super's perspective, the test has been remarkably effective since it was introduced for MySuper options in July 2021 and for Trustee-directed options two years later, but we do not regard the test as being sustainable over the long term in its current form.

We have tried to apply a public policy lens to our submission.

Summary

We believe simplicity should be a vital aspect of the test. Our experience suggests that the more complex any test is, the more likely it can be gamed in unexpected ways and result in unintended consequences. No test will be perfect, and perfection should not be a goal of this consultation.

The overarching objective of the Your Future Your Super performance test is to improve retirement outcomes over the medium to long term. The current performance test approach has a number of issues that are likely to materially impede its effectiveness.

- Failing the test can have dramatic implications for the members of a superannuation fund. Reflecting this, the determination of failure should ideally be based on a holistic assessment of the product and provider. That is currently not the case. Pass or fail is determined by a single metric that is a very partial measure of suitability.
- The test excludes the return related to the product's strategic asset allocation (SAA). It is not clear why this should be excluded as it is often a large part of the return and the SAA is a very important part of portfolio design:
    - For example, a MySuper product could target a similar level of overall portfolio risk to other products through a higher SAA weighting to equities but with lower beta equity exposure. The test would result in a structural component of underperformance for such a product because it does not take a holistic approach (ie it captures the impact of the lower beta equity exposure versus the respective YFYS benchmarks but not the SAA design).
    - Excluding the SAA return also creates an incentive for superannuation funds to "game" their SAA through selecting a low returning set of asset class

exposures (eg including a high cash weighting) to the potential detriment of investment outcomes.

- The test is backward looking and does not consider the "quality" of the organisation providing the product. Ironically, members are generally told that past performance may not be a good indicator of future performance, except in this instance where the performance test framework is based on past performance being a good indicator of future performance!
- The probability of a product failing the test, even if it is well designed, is not immaterial. High-level modelling suggests that the probability for a well-designed product over a given eight-year period could be in the 5% to 20% range, with this probability rising with successive eight-year periods. Moving to a ten-year test period improves the likelihood of avoiding failure - but it remains significant.
  - Over time, there are likely to be increasing numbers of higher quality products that fail the test, just due to the vagaries of markets. This would be to the detriment of the depth and quality of the industry.
  - As the implications of failure of the test may be profound for a superannuation fund's members, there is an incentive for a superannuation fund to take less active risk (versus the YFYS benchmarks) to reduce the chance of failure, potentially to the detriment of long-term performance.
- The test discourages investment innovation. Application of the performance test requires a wide range of benchmarks to measure performance. For traditional asset classes such as Australian equites, reasonable benchmarks are available. However, for some unlisted asset classes and niche strategies, well-aligned benchmarks are unavailable, resulting in higher basis risk or the basis risk is high for some unlisted strategies that have reasonable benchmarks (eg infrastructure). This leads to less incentive to invest in these alternative strategies, potentially to the detriment of investment outcomes.

Alternative approach

We believe that an alternative approach would provide better retirement outcomes. This approach involves:

- Each year, products that underperform by more than the median of a peer cohort by say 0.5% pa would fail that portion of the test. As is currently the case, products that fail the test two years in a row would be closed to new members. This is based on the requirement for a deterministic test, but this is not our preference.
- Our preference is that products that fail the test are reviewed by APRA, which would also include a detailed review of the product provider to determine if its quality is sufficiently high. Some aspects of the current heatmap may be useful for this purpose. If quality is not sufficiently high, APRA would indicate the areas that need to be addressed and review the organisation again in a year's time. If unsuccessful, APRA would consider a range of solutions including closing the fund to contributions and requiring merger. A qualitative assessment is important as no prescriptive test will appropriately cover all circumstances.
  - Governance is a key focus of APRA in this alternative model. In that context, it would be critical that the regulator has the appropriate culture and resourcing.
- The cohort for each product would be determined by APRA and would be based on the key attributes of the product (particularly risk). For example, the MySuper products would form one cohort.
- We are happy to discuss with Treasury how "quality" might be measured in this context. However, we understand that a holisitic assessment is not in scope.

The main advantages of the proposed approach would be better retirement outcomes as:

- The quality of the suite of products available to consumers should be higher due to the:
    - Better basis for determining product suitability through the total return of the product being captured in the performance test.
    - Enhanced quality of superannuation funds reflecting:
        - Competitive pressures as a result of transparency of relative long-term returns
        - APRA assessments of underperforming funds encouraging funds to improve quality
        - Pressure on low quality superannuation funds to merge.
    - More optimal appetite for building strong portfolios for members including active risk taking by high quality super funds because the implications of poor performance are lessened through the APRA quality assessment and because the benchmark used is less knowable.
    - Greater incentive for investment innovation.
    - Reduced complexity allowing a greater focus on achieving investment objectives and performing better than peers.

Consultation questions

2. *Is assessing the implementation of a strategy, as opposed to assessing the choice of strategy itself, a strength or weakness of the current framework?*

We think to answer this question you have to consider MySuper products and other Trustee determined products (TDPs) separately. For MySuper options, the Trustee chooses the strategy for the member and as such it should be included in the test. Part of the Trustee's role here is to choose a good strategy, and the test should not incentivise them to choose a poorly performing one as the current framework does.

For choice products, members have chosen the strategy, even if it is under advice. As such, implementation of the strategy itself should carry greater weight and the current approach is more defendable. One could argue that with the member making the choice, the responsibility should be on them to assess its quality, not a regulatory test. While we have some sympathy for this argument, from a public policy perspective, there are simply too many members making choice decisions, in many cases without the necessary knowledge or resources to assess their ongoing appropriateness, so we favour a performance test for these options.

3. *Can the existing methodology be materially improved, such as by further calibrating benchmarks, to largely address unintended consequences? How could these improvements overcome the incentive to benchmark hug, and remove barriers to invest in emerging asset classes?*

With the lens of simplicity being a virtue, we strongly argue against further calibrating benchmarks and instead focus on holistic outcomes. Our preference would be to remove benchmarks altogether. Naming and utilising benchmarks sold by private companies has led to a material increase in the cost of these benchmarks, to the detriment of members' best financial interests. Furthermore, adding benchmarks will not ease the pressure to add even more later – indeed if anything, it will add to it as those special interests who missed out redouble their efforts. The addition of further benchmarks could end up distorting behaviour further rather than reducing it.

4. *What asset classes do you consider require better coverage in the test? What asset classes are covered well by the existing test?*

The current test provides adequate coverage of the investment universe. Should there be sufficient conviction in an asset class, investors will invest whether the asset class is in the benchmark or not. It could be argued that private equity is particularly disadvantaged by the test. However, if private equity delivers on its promise then it would more than compensate for its off benchmark status. We also note that it is not a large asset class, particularly in Australia.

Our preferred approach is to have a peer-based test with no benchmarks. A peer-based test is less open to herding issues - although this is never absent, even without a test. Because funds have imperfect information about what others are doing there is more latitude and requirement to take risk. The test doesn't need to result in the lowest x% failing. For the performance part it could be that those who underperform the median by say 0.5% pa or more over ten years fail.

5. *Do you consider additional indices covering additional asset classes should be added to the test? If so, please provide the following details for* ***each*** *of your recommendations:*
    a. *Description of asset class*
    b. *Name of recommended index covering the above asset class, including the length of time data is available on the index*
    c. *Details of appropriate fee and tax assumptions for such an asset class*
    d. *Explanation of why you consider this index is appropriate for inclusion*

We are strongly against adding more benchmarks to the test as this will result in increased costs to members - funds purchase these benchmarks at whatever prices index providers think they can get away with.

6. *How should the test cater for new asset classes in the future?*

The test should not cater for new asset classes. Any such new asset class should be able to earn its way in portfolios versus other opportunities. Should investment in such an asset class (for example aircraft leasing, catastrophe bonds, royalties) become material across the industry then at that time a review should be conducted to consider whether a benchmark should or could be added. In our favoured approach, the question would not arise as the weighting would already be effectively weighted in the peer benchmark.

7. *Should the threshold for failure be recalibrated for some products? What evidence supports the need for a different threshold? How could a different threshold deliver better long term returns to members?*

Calibrating the threshold for failure is a very important issue given the current ramifications of it being breached. Under the current framework of a single threshold, riskier products generally have a higher probability of failure and there is a higher probability of incorrectly identifying a product as inappropriate. There is an argument for better aligning the threshold for failure to the risk of the product, with riskier products subject to a higher threshold. If this is done, we suggest that the number of thresholds is limited to two or three to maintain simplicity.

8. *Would retaining the current framework but moving to a simpler structure, such as a simple-reference portfolio of only bonds and equities, address some of the concerns with the current test?*

The move to a simpler bonds/equities framework may increase pressure to reduce allocations to unlisted asset classes and new ideas. It creates a new issue in determining how complex options should be mapped across to the appropriate equity/bond benchmark. Further, there is wide range of risk taken across different equity and bond products. Strategies that might allocate more to, for example, international equities but invest in lower beta strategies would continue to be disadvantaged as they are under the current test. However, we believe that this approach is preferable to the existing approach and it resolves the problem of having more and more benchmarks. It would also be required for a peer benchmark approach to group different products into appropriate cohorts for choice options.

9. *Would the Sharpe ratio be a more appropriate testing approach than the current framework? Would this lead to better member outcomes?*

The Sharpe ratio is not an appropriate measure for the test. It does not accurately reflect risk in unlisted assets. The Sharpe ratio is a poor measure for determining whether retirement incomes will be adequate. Absolute returns matter, and funds should be willing to trade some additional risk for expected stronger long term returns.

10. *How should the benchmark for performance be calibrated?*

We do not believe the Sharpe ratio is an appropriate benchmark.

11. *What data should be used to estimate the Sharpe ratio, and how frequently*

We do not believe the Sharpe ratio is an appropriate benchmark.

Options 2a and 2b

12. *Are either of these approaches better than the existing test methodology (Option 1) or a simple Sharpe ratio (Option 2a)? Are there any other considerations that make this a better or worse option?*

For Option 2b – what constitutes growth? It would prove very difficult in practice to prevent gaming under this option; even within the one asset class risk can be very different for different investments. The actual risk in any investment could be quite different from the assumed asset class risk. For example, a portfolio of high P/E technology stocks would be significantly more risky and growth focused than a portfolio of modestly priced supercore infrastructure stocks - but both would be classified as listed stocks and growth assets.

2c – The issue here is what constitutes risk? Most investors would argue volatility of returns probably isn't appropriate for the test. This test would also likely see more concentration of investments in equities and bonds.

13. *Are there any other alternative single-metrics that would be superior in addressing the principles set out in this paper? How would they provide a better testing framework? What net benefits do they provide over other proposed metrics?*

A benchmark of the average of peers would be superior. It requires no external benchmarks, and it is harder to game due to a lack of information as to what peers are doing real time. For MySuper options we suggest this test as described is sufficient. At times, it will encourage greater, but not imprudent, risk taking. Currently funds take a keen interest in their position in MySuper surveys. We don't think the change to a peer benchmark would significantly alter how they invest MySuper products given the competitive tension that exists. This approach would mean that strategic asset allocations would be implicitly included in the test. In addition, it would be measuring against what members actually get - returns, not risk adjusted returns. One criticism of this test has been that the bottom x% of options would fail each year. This does not necessarily have to be the case. It could be that those who underperform the median option over ten years by more than say 0.5% pa fail. This would be similar to the current test in that respect.

Our observation on risk-adjusted returns is that such an approach would favour well diversified portfolios but not absolute returns. To minimise the chances of failing the test, investors are likely to try and move toward a more equally weighted contribution to risk from each prescribed asset class. This would mean more cash and bonds and lower long-term returns for members – leading to poorer retirement outcomes.

For choice options, funds could choose from (for example) four different growth/defensive categories (up to 20%, 40%,60% and 80% growth allocations). APRA would need to define the growth allocations for different asset classes and decide how to treat new asset classes. This would be a similar exercise as is required for the simple bonds/equity benchmark option. Of course, this would be open to potential manipulation by funds - but that is also the case with the current test. It would likely add incentive for funds to position their options towards the top of the growth range. This would likely be greater than would be the case for MySuper options as our experience is that there is less competitive tension in these non-MySuper options currently. Despite these flaws, we think this approach would be better than the current test.

14. *What incentives would these alternative single-metric options provide trustees, and what would be the consequence of this for member outcomes?*

With a peer test, funds would have to make decisions based on more imperfect information than is currently the case. Paradoxically, that would make funds more open to new ideas to try and gain some competitive advantage. Funds that are close to failing the test are likely (depending on the environment) to look to increase the risk they take. This risk exists with the current test in that failing funds are likely to take materially more risk than is set out in their SAAs in trying to escape perdition.

15. *Would greater alignment to the APRA heatmaps improve the sophistication of the test?*

Is sophistication a goal of the test? As noted in our introduction we believe complexity lends itself to unintended and potentially misunderstood consequences. The APRA heatmaps do provide a more holistic assessment of the sustainability of funds.

16. *Would it reduce incentives to benchmark hug and improve member outcomes?*

Using five-year periods would make funds cautious about risk sizing. It is likely to result in more ‘growthy’ SAAs to some extent but if anything, smaller positions against the simple reference portfolio. We think the investment horizon in the performance test of ten years is about right. We have commented on the issues and benefits with the SRP above.

17. *Is correlation between metrics an issue? If so, how should this be addressed?*

Correlation between benchmarks isn’t necessarily an issue. Given a principle of simplicity we would prefer to see fewer tests (one net cash inflow test, one performance test) rather than several of each. Multiple tests would make gaming harder, would increase failures and would likely result in funds trying to reduce active risk to the extent they can.

18. *Should the test capture all the metrics in the heatmap? If not, what metrics?*

This is answered above.

19. *How would the benchmark for performance be calibrated for chosen metrics? How would these metrics combine to determine overall pass/failure of the test?*

With one performance test of each type (performance, cashflow and fees) we would advocate simplicity and weight these equally based on a simple pass or fail.

21. *Would this framework improve the sophistication of the test? Would it reduce incentives to hug benchmarks and improve member outcomes?*

We don’t see “sophistication” as being an appropriate goal of the test. The incentives to hug benchmarks would be undiminished, although there would need to be some trade-offs. Risk-adjusted targets are not appropriate. Risk is not clearly defined and what is risky over one year may be least risky over 20 years and vice versa. Risk adjusted approaches are likely to lead to more diversification but at the expense of long-term returns and therefore member retirement balances. They also incentivise investment in unlisted asset classes and increased liquidity risk.

Our experience on CPI + x benchmarks is that these are a signal to members of what they can expect from investing in an option over the longer term. It is very difficult for funds to invest directly into CPI + investments outside of inflation linked bonds, and perhaps infrastructure assets with CPI linked revenues. We would not recommend using CPI + targets as part of the test.

22. *Would this approach be more, or less, favourable than the heatmap approach?*

On balance, less favourable.

23. *What would the costs of implementing this approach be? What would the benefits be?*

We do not believe this approach should be implemented.

24. *Are these the right measures of performance or are there other more important indicators of performance that should be measured in addition to or instead of those outlined? What metric should be used to assess these indicators?*

There are no perfect measures for performance; no objective test that would work well in all circumstances. On the investment performance side there is a large amount of luck involved in product performance and this can overwhelm any skill even for the best run of funds. There is no substitute for a well-resourced regulator with the ability to examine funds for suitability and to make sound judgements.

We note that there is no assessment of liquidity risk in the test and this, in some environments, could prove to be very important.

26. *How would an alternative framework be constructed according to the elements outlined above? Please provide specific details.*

We have provided our view on an appropriate investment performance element above.

27. *How would this framework more effectively advance the principles outlined in this paper?*

Our arguments on this are also outlined above. In summary, because peer benchmarks are more unknowable funds are forced to take risks against them. They also allow for innovative new ideas to permeate the industry, albeit perhaps more slowly than with no test.

37. *Should fees be measured at the current option level, or should they be measured on a different level? How would this be achieved?*

To avoid adding further complexity we believe the fees should continue to be measured at option level.

38. *Are the current assumptions made in comparing fees acceptable? For example, should the $50,000 representative member balance be adjusted based on the median member balance for a product cohort?*

We believe appropriate consideration needs to be given to the fees charged to member accounts. As the industry has matured so have member balances. We believe comparison or use of the representative member balance of $50,000 no longer remains appropriate and alignment with a more relevant balance should be considered.

We do not think adding the complexity of using the average balance of each fund is needed as it removes some ability for comparison across the test. Our suggestion would be to align with existing cohorts available within the Heatmap: the $100,000 balance for accumulation accounts and the $250,000 for pension phase assessment, while also acknowledging this will need to shift over time.

39. *Is a peer comparison of fees the best way to measure fees? Is there a better approach to benchmarking fees? If so, how should this work?*

Consistent with our earlier responses, our preferred approach is to have a peer-based test with no benchmarks other than the median.

40. *What product cohorts should be considered? How should different cohorts be defined where products could meet multiple cohort definitions, such as single-sector retirement products?*

The principle of simplicity should be maintained, MySuper for both single investment strategy and lifecycle stage cohorts should be considered, leveraging currently available data, along with Trustee directed and retirement products. Fees should be measured at option level as with cohort definition a level higher at product enabling differentiation of accumulation and retirement products.

41. *How many years of fees data is appropriate to test? Should a greater weighting be given to certain years?*

We are aligned with commentary within the consultation paper regarding the test period for the RAFE to be assessed over the entire lookback period with a greater weight being given to the current year of fees. We would not like to see an additional reporting burden placed on Trustees (which we note ultimately increases costs for members) so would suggest the look back period only extend to the data currently available and extend over time.

43. *How should the consequences be amended to better account for edge cases or different cohorts that fail the test for reasons beyond the trustee's control?*

It may be that the consequences should not be amended. If the price for being able to better account is a significant increase in complexity, it may introduce unintended consequences elsewhere.

44. *How could these provisions be effectively ring-fenced so that it applies only to the edge cases and not failures at large?*

See above.

47. *Are there any key barriers to consolidating closed and underperforming products? What quantitative evidence is there of these barriers? How do these weigh against other reasons a person may choose to remain in a product?*

Outside of MySuper options, the member's right to choose is an issue. While any test may result in failure, it may well be that the member is getting what they set out to achieve. Particularly for more defensive options, return is often not the primary motivation. It may be for example that the member wishes for positive returns each and every year and that that is more important to them than long term returns against any benchmark. For any product, there will likely be some members who wish to remain in it. This makes closure more difficult without compulsion.

48. *What evidence do trustees use to demonstrate that remaining in a closed and underperforming product is in the best financial interests of members, compared to moving to a performing product?*

The issue here is what is a "performing product"? APRA cannot determine what these are. Some of the best performing MySuper products from a decade ago became some of the worst performing over the subsequent decade. Funds' ability to perform changes over time, and luck plays an important role. Any test must take this into account, especially when it comes to pointing members toward alternatives.

50. *Should APRA receive increased regulatory powers to direct superannuation trustees to consolidate underperforming products?*

Yes it should, and here the sustainability of the product should be given greater weight then performance alone. Performance can turn on a dime but a small product bleeding cashflows is much more difficult to turn around.

Any questions or clarifications about this submission can be addressed to:

Rebekka Power
Head of Communications
rpower@visionsuper.com.au

Yours sincerely

Michael Wyrsch
Chief Investments Officer