19 April 2024

Superannuation Efficiency and Performance Unit
Retirement, Advice and Investment Division
The Treasury
Langton Crescent
Parkes ACT 2600

By email: YFYS@treasury.gov.au

Dear Sir/Madam

Subject: Submission – Annual Superannuation Performance Test – design options

We are pleased to provide this submission on the Government's review of the Annual Superannuation Performance Test.

WTW is a leading global advisory, broking and solutions company that helps clients around the world turn risk into a path for growth. With roots dating to 1828, WTW has colleagues serving more than 140 countries and markets. We design and deliver solutions that manage risk, optimise benefits, cultivate talent and expand the power of capital to protect and strengthen institutions and individuals. In Australia, we provide actuarial, communication, technology and superannuation and investment services to a broad range of defined benefit and accumulation superannuation funds including standalone corporate funds, industry funds, master trusts and master trust sub-funds.

We set out below our high level comments and views.

We support the four principles proposed

We agree with the four principles proposed, though we believe that improving member outcomes should be the primary factor – with the implication being that the test should not discourage investments that are in the best interests of members or unreasonably restrict member choice.

While we do not propose that it be a further principle, we note that any change to the performance test should also be assessed against the criteria of not increasing systemic risk. A performance test regime that (further) encourages herding behaviour by funds would only add to the existing risk of coordinated losses to members in the event of a materially adverse market event, potentially leading to a loss of confidence in the wider system.

Nick Callil FIAA
Senior Director, Retirement

Level 32, 385 Bourke Street
Melbourne
VIC 3000
Australia

T +61 3 8681 9800
D +61 3 9655 5163
M +61 407 662 464

E nick.callil@wtwco.com
W wtwco.com

Towers Watson Australia Pty Ltd
ABN 45 002 415 349 AFSL 229921

Any single metric will have flaws

Any single past performance-based metric will have flaws. These flaws will include the well-known qualifier that past performance is not a reliable indicator of future performance. More concerningly, however, a single metric will necessarily provide a clear incentive for funds (whether consciously or implicitly) to manage their investment program and product design to ensure a favourable outcome in regard to that metric (or, at least, to minimise the likelihood of an adverse outcome). While this might be seen as benign, it can work against the overall long-term interests of members where the outcomes captured by the single metric do not align perfectly with overall member outcomes. The current test, for example, which measures implementation relative to an (implied or actual) asset allocation benchmark, potentially incentivises funds to invest in listed assets with easily investible benchmarks. However, any single metric test will necessarily create the potential for distortion of the decisions around investment design or implementation, to the possible detriment of overall member outcomes. The potential for such distortion is high given the outcome of failure of the test is so grave.

A multi metric test is a better structure

In our view, therefore, a better structure for the performance test would be one which reduces the reliance on a single metric. We see the advantages of a multi metric testing structure as twofold:

- a multi metric test provides a more diffuse testing structure, removing the 'spotlight' effect of a single metric, and hence the capacity / incentive for funds to focus on a single metric. Provided the elements of a multi metric test are well targeted, this is more likely to encourage funds to focus on broader member outcomes.
- Introducing more than one 'leg' to the test allows a broader range of the components of overall fund performance to be incorporated in the test. As noted above, the current test measures implementation relative to a product's (actual or implied) asset allocation benchmark only – a very limited component of the product's overall performance. A multi metric approach allows other components – such as asset class selection – to be incorporated into the test, with the aim of increasing alignment with overall member outcomes. (We discuss below the possible forms of the multi metric test).

While the imperfect nature of any past performance based metric is acknowledged, a multi metric test can better allow for this via modifying the criteria for failure (e.g. failure occurs where two out of two, or two out of three metrics are failed).

Multi metric test discussion

We propose that the design of a multi metric test be subject to further consultation. This could include a more focused examination of elements such as:

- which of the proposed metrics, and in what combination, best realises the structural advantages of a multi metric test as discussed above
- how the threshold for failure (currently 50bps) should be applied to the new tests.

Our preliminary view is that a two or three metric test, with one of the metrics being either the current test or an improved version of the current test (see below), is likely to be most appropriate. The other tests could be a peer comparison of risk adjusted returns (Option 2b) and risk adjusted returns relative to a Simple Reference Portfolio frontier (Option 2c). Failure would occur when two (or more) of the tests are failed over a relevant measurement period.

The reasons for suggesting such an approach are that:

- adding the additional metrics improves the testing regime by assessing performance in different ways (including explicit allowance for realised risk adjusted returns); and
- the 'two out of three' fail mechanism dilutes the inherent weakness of each metric (such as the difficulty in measuring growth and defensive, and the discrepancies in measurement of volatility between listed and unlisted assets);

We also believe that the current test (whether used as a stand alone test, or as a component of a multi metric test as we have suggested) can be improved by adjusting the failure threshold (currently 0.5%) in line with the product's relative risk level (tracking error). The Appendix set out details of this suggestion, including how it would address the various concerns with the current test raised by stakeholders.

Cost of replacing current test

We acknowledge there is a cost to replacing the current test. Funds have internalised the current test and set up processes to manage and monitor performance against this test. However we do not believe that this cost should preclude adopting a new, better test. If a test can be designed that better aligns the incentives for funds with long term member outcomes, the positive impact on consumers will likely exceed any cost of transition.

Transitional issues should be addressed through further consultation to ensure that there is no unreasonable imposition of new criteria retrospectively.

We note that a multi metric test wouldn't necessarily require more data to be collected, as the data required to implement the additional tests are in many cases the same data required for the current test, limiting any additional burden on funds in terms of reporting and on APRA in terms of collection.

Scope of the Test

We propose that the test should not apply to retirement products. The retirement phase is going through an important phase of development at the moment, with a number of funds seeking to enhance their retirement offerings (including introducing new retirement products). Given the different characteristics of such products (including longevity protection in some cases), return-based tests are ill-suited to measuring their appropriateness or effectiveness. Further, given the stage of development of the retirement segment, we believe an imposition of a test could serve to inhibit innovation.

Consequences of Failure

We have previously argued (https://treasury.gov.au/sites/default/files/2021-02/c2020-124304_willis_towers_watson.docx) that exercise of discretion by APRA before imposition of the most serious consequences of the test should be considered. We would continue to recommend this approach – particularly in the case of Trustee directed products, where (unlike most MySuper product members) consumers have explicitly chosen to invest in the product.

We would be pleased to discuss this submission with you or to provide any further information required.

Yours sincerely

Nick Callil
Head of Retirement Solutions, Australia

Jonathan Grigg
Director, Investments

Appendix – modified current performance test

The primary shortcoming of the current performance test as an *implementation test* is that it discourages taking risk relative to the current set of benchmarks – an investment might be in the best interests of members but result in relative risk that increases the risk of a superannuation product failing the performance test.

This is because the test seeks to (imperfectly) adjust for overall investment risk, but does not adjust for relative risk (the amount of active risk taken). This is reflected in the concerns raised by stakeholders – particularly that the current test encourages short-term decision making, incentivises benchmark hugging, results in a lack of investment flexibility, reduces choice and discourages both diversification and active management.

We believe that the current test can be improved by adjusting the 0.5% threshold for failure in line with the risk taken relative to the product's SAA benchmark.

As an example, adjusting the test based on realised tracking error to the SAA benchmark could result in failing margins of:

- 0.25% p.a. for a product with tracking error of 1% p.a.
- 0.50% p.a. for a product with tracking error of 2% p.a.
- 0.75% p.a. for a product with tracking error of 3% p.a.

We summarise below how this modification would address the concerns raised by stakeholders. We also view this modification would result in a test that is more grounded in statistical science and more accurate in identifying poorly performing funds.

Concerns raised by stakeholders	Addressed/Mitigated by adjusting for relative risk?
Focusing on investment implementation over other measures of performance	**Somewhat** – allows for the assessment of moves away from SAA benchmark. ***We suggest a multi-metric framework that includes a separate metric(s) less focused on implementation***
Encourages short-term decision making	**Yes** – because the test would adjust for the short-term uncertainty of investing in assets that differ from underlying benchmarks, it would significantly reduce the push towards short-term decision making. It would also reduce the incentives for funds to adopt a short-term approach in order to prioritise passing the test over member outcomes
Incentive to hug benchmarks	**Yes** – because this metric would adjust for benchmark-relative risk, the current incentive to hug benchmarks would be removed
Lack of investment flexibility	**Yes** – adjusting for relative risk would take away the disincentive to invest in assets that are not well represented in the benchmarks, including emerging asset classes and values-based investments
Reduced choice, diversification, and active management	**Yes** – assets that are not well represented in the benchmark indices come with higher tracking error, which would widen the potential relative outcomes that are allowable without failing the test. Similarly, adjusting for relative risk would avoid the current disincentive for active management