Associate Professor Geoff Warren
Research School of Finance, Actuarial Studies and Statistics
College of Business and Economics, Building #26C
Kingsley Street, Canberra, ACT, 2600

Mobile: [redacted]
Email: [redacted]

14 October, 2022

Attention:
Director of Superannuation, Efficiency and Performance Unit
The Treasury
YFYS@treasury.gov.au

Submission to the Your Future, Your Super Review

Dear Review Team,

I welcome the opportunity to provide input into the Your Future, Your Super (YFYS) review. The views offered in this submission are my own, and should not be taken as representing the official position of the Australian National University.

I do not intend to directly address the questions raised in the Consultation Paper dated 7 September 2022. Rather, I aim to draw on my own expertise and experience by offering selected insights that I believe may be of value. My comments specifically relate to the broad approach used to evaluate superannuation fund performance, rather than the finer specifics of the YFYS performance test ('YFYS test'). My research over the years has touched on performance evaluation from multiple perspectives; and I provided input to the Productivity Commission on the design of their performance test under their 2018 review into the efficiency and competiveness of the superannuation industry ('PC review').

My submission largely relates to the 'unintended consequences' element of the YFYS test. I focus on how the broad design of the YFYS test may lead to poorer outcomes for superannuation fund members and how they might be addressed, although I understand that change to the nature of the test are unlikely to occur off the back of this particular review. The two main adjustments I suggest are: (a) broadening the range of performance measures considered, and (b) introducing an 'informed reviewer' rather than relying on a 'bright lines' test, probably involving poorly performing funds being required to show cause why they should not be declared 'underperforming'.

I do not intend to comment on evidence of the YFYS test leading to perverse behaviours in practice, design of the YFYS benchmarks, application to choice products, or the YourSuper comparison tool. I see others as better placed to comment on these elements, on which I am sure you will get ample input.

My contact details can be found above if you wish to discuss my submission.

Yours sincerely

G Warren

Dr Geoff Warren

The YFYS test has yielded some benefits for fund members

Assessment of superannuation funds under the YFYS test has yielded some benefits. The PC review highlighted a meaningful ‘tail’ of superannuation funds that had underperformed substantially over an extended period. The YFYS test has acted as a catalyst to help ‘clean up the tail’ through many of these underperformers exiting the industry. The YFYS test has placed pressure on funds to reduce fees, and has probably enhanced confidence in the superannuation system. These specific developments are likely to leave members better off.

Role of performance evaluation – purposes and implications

It is worthwhile to consider the role of evaluating superannuation fund performance for context. My key point is that performance evaluation has multiple purposes and various implications. Nevertheless, the YFYS test have been largely directed towards one purpose over others, which might be summarised as ‘holding poorly performing funds to account’. Arguably the accountability mindset has produced a test that glosses over other important aspects. Four roles for performance evaluation include the following:

a) *Acknowledging success, and identifying poor practices* – The traditional purpose in undertaking performance evaluation is to gauge what has worked so it can be rewarded, and what has not worked so that remedial action may be taken. The YFYS test has been largely focused towards the latter. The use of evaluations to recognise and reinforce beneficial practices has been less prominent, although the YourSuper comparison tool is partly directed towards highlighting funds that have been more successful.

b) *Incentivise appropriate behaviour* – An often-overlooked but important reason for – and implication of – undertaking performance evaluation is that it incentivises behaviours. This is best summed up in the Peter Drucker adage “what gets measured gets managed”. This dimension was given minimal attention in designing the YFYS test. Many of objections over the YFYS test relate the creation of perverse incentives for superannuation funds to adopt behaviours that may worsen member outcomes – concerns that seem to have been brushed aside so far. Given that the ultimate aim of the YFYS test is to improve member outcomes, it seems amiss to ignore the potential consequences of the test itself for behaviours and hence also member outcomes.

c) *Building member confidence* – Comfort that funds are being monitored to help guard against poor practices can enhace confidence in the superannuation system at large. The YFYS test appears to have fulfilled this purpose, noting that the formalisation of fund performance evaluation appears almost universally received as a positive development.

d) *Encouraging competition and continual improvement* – An evaluation process can help encourage competition and spur continual improvement, especially if the results from the evaluations are made public and guidance on how to improve is provided. The extent to which the YFYS test has been effective in this regard is debatable, given it has largely spurred activities that are primarily designed to avoid failing the test rather than being focused squarely on outperforming competitors or improving member outcomes. While fee reductions are one notable improvement that has arisen from steps taken to avoid failing the YFYS test, it is not clear that the overall momentum of change has been positive for members.

In summary, the primary focus under the YFYS test has been holding poor performers to account. Meanwhile, little consideration has been given to the behaviours that are being incentivised and whether they are appropriate. The latter is important and deserves greater consideration.

A further consideration is that performance evaluation is typically managed and often gamed. Various accounts suggest that many fund trustees are now focused on managing the YFYS test as they strive to minimise the chance of failing the test given the dire consequences. One likely implication of these efforts will be a reduction in the effectiveness of the YFYS test in identifying poorly-managed funds, as funds pull out all stops to avoid a test failure. Meanwhile, the adverse impacts for members from perverse incentives are only likely to build as fund adjust their behaviours. Looking forward, the YFYS test is hence likely to become less effective in holding poorly managed funds to account as the damage builds from perversive incentives, thus potentially tilting towards being a net negative for members.

The broad design of the YFYS test is the main problem

There are two overarching problems with the broad design of the current YFYS test. First is its formulation as a 'bright lines' test with severe consequences for failure: this sits at the foundation of many of the perverse incentives. The second is that it is a backward-looking test based on imperfect science. The latter undermines the effectiveness of the YFYS test in identifying genuinely 'poor funds' and hence in performing its member protection role.

Design problem #1: Bright lines test with severe consequences for failure

The severe consequences for failing the YFYS test create an overbearing incentive for fund trustees to ensure they do not fail. Managing the test outcomes thus becomes a first priority. This would not be an issue if the YFYS test was well-aligned with member outcomes. However, this only occurs along the limited dimension of how effectively trustees have implemented the stated strategic asset allocation that the member has accepted. The consequence is that trustees may be incentivised to act in ways that are not necessarily in the best interests of their members, as other objectives and services that may matter to members become subsidiary the YFYS test, e.g. pursuing a real return objective, working towards adequate income in retirement, limiting risk to total returns, improving member services, etc.

Another issue is that the YFYS test may induce herding around anointed benchmarks to limit the risk of test failure, with the implication that the test will tend to direct how the industry at large invests. Also, tracking error becomes a primary risk measure regardless of its relevance for members. Potential impacts include reluctance to diversify into assets that are not covered by the benchmarks, stifling of innovation, and potentially the introduction of systemic risks through common portfolio structures.

Design problem #2: Backward-looking test based on imperfect science

The ultimate aim is to identify funds that are likely to be poor performers in future. There is an implicit assumption underlying the YFYS test that poor funds can be identified using past returns alone. Unfortunately, the science of performance evaluation is not up to the task of identifying a fund that is likely to continue underperforming with high confidence.[1] Beyond the overarching point that past performance is an unreliable predictor of future performance, the efficacy of the test is also subject to technical issues such as the suitability of the benchmarks, lack of adjustment for risk, whether strategic asset weights are being set in an appropriate manner, and data quality.

A major shortcoming of historic performance is that it is an unreliable indicator of 'skill'. Funds that fail the test may have made sensible decisions in the circumstances with member interests at the forefront, but the decisions did not happen to work out at the time. A notable example is the funds failing the YFYS test in the first round due to adopting defensive positions that have subsequently paid-off. The positions taken may have been sensible and prudent, but the funds failed the test and suffered the consequences nevertheless. Equally, poorly managed funds may make bad decisions or not operate in the best interests of members and still pass the test. Perhaps more importantly, a solely historical test gives no consideration to changes that may have addressed the issues that led to poor performance, such as replacement of management or renewal of the trustee board.[2] The only way to evaluate these aspects is to examine the basis of decisions to gauge if they were appropriate in the circumstances, and to analyse fund capabilities such as the quality of management and processes

[1] Some reasons include: changes in management or processes; structural shifts in market conditions (some investment managers or processes are suited to certain environments); mean reversion in markets, such that previously losing positions become winning positions; and random variation in uncertain markets, i.e. bad luck.
[2] Fund mergers also muddy the waters.

Addressing the main problems

While there is room to improve some aspects of how the YFYS test is implemented, such as the benchmarks being used or perhaps incorporating some consideration for risk[3], this would be playing around the edges. The two main problems raised above are far weightier, and can only be addressed by significant adjustment of the test design. My suggestion is two-pronged: *(a) broaden the set of performance measures*, and *(b) incorporate interpretation by an 'informed reviewer' prior to declaring a fund as "underperforming"*. These suggestions are jointly aimed at diluting the desire to manage to the YFYS test and its benchmarks, thus limiting the perverse incentives that are created by an uncompromising test.

Broadening the set of performance measures

Broadening the set of performance measures would reduce reliance on a single, imperfect performance measure. It would also dull the influence of the chosen benchmarks on behaviour. Broadening the range of measures might emulate what APRA was originally trying to achieve with its heatmap, perhaps with selected risk measures being added. One consequence of broadening the set of performance measures is that they would need to be balanced against each other, which may require an element of interpretation. Thus this first suggestion and the second suggestion below are linked.

Interpretation by an informed reviewer

Introducing an informed reviewer into the evaluation process would have two main benefits. The first is it would facilitate undertaking the evaluation through a forward-looking rather than backward-looking lens. The central question that the reviewer would address is whether poor performance is likely to continue into the future. The second is that fund trustees would be provided with confidence to pursue activities that are in the best interest of members without the fear that they will necessarily be branded a poorly performing fund based on performance measures alone. This confidence would arise from trustees knowing they will have an opportunity to demonstrate that they were doing the right thing by their members.

Reviews could occur on a 'show cause' basis. That is, a fund that has underperformed based on the performance measures would need to provide clear evidence that poor performance is unlikely to persist to avoid being branded 'underperforming'. Trustees would be required to demonstrate that their decisions were sensible and aligned with member interests, and/or that the problems have been addressed that led to poor performance in the past.

Reviews might be undertaken by a specialised division within APRA, or perhaps an entity that is created for the purpose, e.g. panel of experts. The process might be comparable to the 'manager research' that is undertaken by asset owners and consultants to gauge if an investment manager is suitable for use. Manager research typically draws on engagement with the manager as well as data analysis, including an examination of the sources of performance and the rationale behind investment decisions. The process could offer scope to interpret performance in light of the fund's 'mandate' by taking the nature of the member base and member preferences into account, e.g. choices around investing sustainably.

---

[3] Incorporating a single risk measure into the YFYS test is problematic. Given that the YFYS test encourages superannuation funds to manage the risk of failing the test and hence tracking error versus the benchmark, the aim would be incorporate a risk measure that takes the member's perspective. The problem is that members are concerned about different risks, and have differing tolerances for those risks depending in their risk appetite and personal circumstances (age in particular). Another complication is that most existing 'technologies' to measure risk are based around return volatility, e.g. standard deviation, perhaps drawdown. Volatility is an inappropriate risk measure when investing over the long run, where risk relates to failure to generate sufficient to wealth to support objectives (e.g. adequate income in retirement) or perhaps the potential for a permanent loss. Further, shifting to less volatile assets can increase the chance of failing to achieve a long-term objective if they offer insufficient returns. A dashboard of risk measures may be more appropriate than a single volatility-based measure.