WHITE PAPER

*Not everything that counts can be counted,
and not everything that can be counted counts.*

— ATTRIBUTED TO ALBERT EINSTEIN

BY **BILL SPITZ**, PRINCIPAL

Perspectives on Performance Measurement and Evaluation

Every investor is appropriately focused on the return on her portfolio, and that concern assumes the mantle of legal responsibility for trustees of fiduciary accounts such as trusts, endowments, foundations, and pension funds. Thankfully, there are numerous vendors that provide return data including publications such as Morningstar, bank custodians, investment consultants, investment managers, and brokerage firms. But, despite reams of data provided by these resources, may I be so bold as to suggest that much of what constitutes performance evaluation is fraught with errors, misplaced focus, hasty and unsupported conclusions, and false precision.

continued on next page ›

And, these problems collectively lead to a large number of unnecessary and potentially expensive decisions. In this paper, I enumerate many of what I deem unsatisfactory practices, most of which are implemented by well-intentioned individuals who are trying to carefully exercise their responsibilities. And, more important, I provide a framework for better performance evaluation with particular emphasis on what the numbers can and cannot tell us. In the final analysis, I hope that you will be convinced that performance evaluation is rather qualitative despite all of the number crunching.

Common but Questionable Practices

I have attended literally hundreds of investment meetings and have found that relatively few committees focus much attention, if any, on what is arguably the most important return datum, the total return on the fund versus the purpose for which it was created. For example, an endowment's return goal is equal to its spending rate plus the rate of inflation, typically in the range of 6-8%. By achieving that return, the fund accomplishes its objective which is to preserve the purchasing power or impact of the endowment in perpetuity. A corollary is that there is typically relatively little discussion of the fund's asset allocation which is the primary driver of returns. Many committees do pay attention to the total fund return versus a blended market benchmark. But, what good is beating market yardsticks if the fund is not accomplishing its basic purpose? Moreover, to the extent that there is discussion of the total fund return, it is often directed primarily toward a comparison with competitive institutions.

Instead of focusing on the big picture, the great majority of committee meetings is typically devoted to examining the return on each component of the fund and its managers in excruciating detail. While it is indeed appropriate to evaluate the individual segments, there is a powerful human tendency to scan the performance page, find a category that is behaving poorly in either absolute or relative terms, and then spend a disproportionate amount of time and energy discussing it. Ironically, the component under the microscope frequently represents only a small percentage of the overall portfolio! A related problem is that a properly diversified portfolio should always have segments that are underperforming in absolute terms because they were included in the portfolio specifically to counteract poor performance in other segments. Simply stated, if a portfolio does not have some laggards, then it is not really diversified and will ultimately experience a painful downdraft. Yet, in the heat of performance evaluation, committee members frequently forget this principle and feel a strong urge to discipline the problem child.

continued on next page ›

The next, and perhaps most destructive behavior, involves making decisions on the basis of inadequate time horizons. First, most Investment Policy Statements call for evaluating performance over a complete market cycle which is frequently defined as three to five years. But, historically, a complete cycle in the equity market has averaged almost exactly ten years. I have never encountered a committee with the patience to evaluate results over a ten year period! However, even ten years is generally not long enough to differentiate between luck and skill. For example, to statistically demonstrate that a large cap stock manager has outperformed the benchmark by 1% per year due to skill requires about 39 years of data, and the time frame is even longer for small cap and international stock managers because they generally have greater tracking error. So, judgements based on a few years of data are largely based on random noise, good or bad luck, the cyclicality of investments styles, and other factors that have little or nothing to do with the presence or lack of skill. Remember, no one selects a fund or hires a manager with a poor long term track record so there was an initial presumption of skill. Is it reasonable to believe that the manager has lost her skill after a couple of bad years?

A fourth problem is that many people casually select benchmarks to evaluate the various components of the fund. This is perfectly understandable given that there are literally thousands of benchmarks available, and many strategies and managers do not fit neatly into a well-defined asset category with a clear-cut benchmark. Nevertheless, it is critical to do the best job possible of selecting the measuring stick. First, most professional investors, whether consciously or not, manage at least to some extent to the benchmark against which they are being evaluated, so the choice of benchmark influences the actual investment of the portfolio. Second, a poor choice can lead to spurious hiring and firing decisions.

Finally, very few people carefully evaluate results in the context of the amount of risk taken. Risk measurement is a technical topic that is beyond the scope of this paper, so suffice it to say that most performance measurement services do the math, they calculate the volatility of the portfolio and the risk adjusted return on its components. Why should you care about overall volatility? The basic math of compounding demonstrates that a more stable portfolio will end up with more money than a more volatile one with the same average return. And, dampening volatility reduces the odds of an emotional, destructive response to market turmoil. But, the real challenge is to evaluate rather than measure risk. Is the overall volatility of the fund consistent with its goals and the level of risk tolerance of the governing body? Are the various segments of the portfolio doing their job in terms of reducing risk through diversification? And, are the riskier components of the portfolio generating enough extra return to earn their keep? None of these are easy questions so most people ultimately throw up their hands and simply focus on raw returns.

continued on next page ›

So What?

What is the impact of these practices that I have characterized as problematic? First, poor performance evaluation frequently results in perverse timing decisions; the data clearly demonstrates a tendency to sell assets after they have underperformed and vice versa. And, this is true of both individual investors and institutions. Similarly, investors change funds and managers too frequently based on short time horizons, random noise, and failure to recall what role each component was designed to play in the portfolio. And, these portfolio changes can be expensive. There are transaction costs involved in changing horses and taxable investors may well incur tax liability. Most important, various studies suggest that the new vehicles or managers chosen experience future returns that are indistinguishable from those of the vehicles that were replaced. So, the net effect is no material improvement in future performance and significant cost and administrative headaches along the way.

How Should One Approach Performance Evaluation?

As you will see, performance measurement and evaluation are ultimately both subjective and qualitative, but the following recommendations should provide a useful starting point:

- **The most important measure of success is the performance of the overall fund versus the purpose for which it was created.**
- **The second area of focus should be the return on the total fund versus a blended benchmark which consists of an appropriately weighted benchmark for each underlying asset class.**
- **Benchmarks should be chosen with care to ensure that they actually correspond to the strategy under consideration.**
- **It is critical to understand and properly evaluate each asset class in terms of its role in the portfolio.**
- **While the responsibility and perceived pressure of governance make it difficult to employ a long time horizon, judgements based on a time frame of much less than ten years tend to be dominated by random factors.**
- **For even longer time frames, it is extremely difficult to tell the difference between skill and luck.**
- **Either quantitatively or qualitatively, an important component of performance measurement is risk assessment.**

continued on next page ›

Limitations

Unfortunately, even if these recommendations are followed to the letter, performance measurement and evaluation is a squishy process. First, with respect to manager and fund evaluation, very few teams and organizations remain unchanged long enough to meet the statistical significance test of skill versus luck. So, manager evaluation ends up being largely qualitative. Therefore potential changes should not be based on even intermediate term return data but factors such as the following:

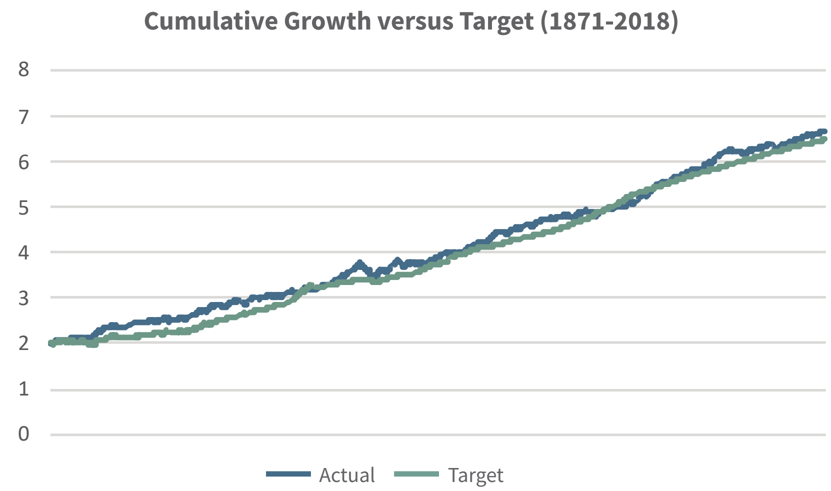
- **High Fees**
- **Changes in key personnel or firm ownership**
- **Addition of new products or strategies**
- **Unchecked growth in assets under management**
- **Perceived changes in investment approach**
- **Performance outside of expected range of variability**
- **Operational issues, poor controls, weak risk management**

None of these can be reduced to a single statistic which confirms the assertion that performance evaluation must be holistic rather than a simple review of reams of data.

Very few teams and organizations remain unchanged long enough to meet the statistical significance test of skill versus luck.

continued on next page ›

Several times, I have emphasized the importance of focusing on the return on the portfolio versus the purpose for which it was created. But, even that analysis ends up being subjective. The following chart depicts the cumulative growth (logarithmic) from 1871-2018 of a 60% US Stock/ 40% US bond portfolio versus a target of inflation plus 5%. It demonstrates that the traditional balanced portfolio has been successful at accomplishing the goal of a typical endowment or foundation.



But, let's break down the picture into ten year rolling returns, again over the period 1871-2018.

Important Notes And Disclosures

This White Paper is being made available for educational purposes only and should not be used for any other purpose. Certain information contained herein concerning economic trends and performance is based on or derived from information provided by independent third-party sources. Diversified Trust Company, Inc. believes that the sources from which such information has been obtained are reliable; however, it cannot guarantee the accuracy of such information and has not independently verified the accuracy or completeness of such information or the assumptions on which such information is based. Opinions expressed in these materials are current only as of the date appearing herein and are subject to change without notice. The information herein is presented for illustration and discussion purposes only and is not intended to be, nor should it be construed as, investment advice or an offer to sell, or a solicitation of an offer to buy securities or any type of description. Nothing in these materials is intended to be tax or legal advice, and clients are urged to consult with their own legal advisors in this regard.

continued on next page ›

As you will note, the actual return differs significantly from the target over long periods. (On both the plus and minus side) In fact, the actual return is outside of a band of .9 to 1.1 times the target return about 85% of ten year periods! So, even assessing the return on the portfolio for that time horizon can provide misleading information regarding long term progress. The message is that reviewing any performance data requires patience and equanimity.

Where Does All of This Leave Us?

I was giving a talk on this subject to a group of trustees of non-profit organizations, and one said: "Bill, I understand that a great deal of performance data does not lead to definitive conclusions. But, as a fiduciary, I feel a responsibility to ***do something*** when things aren't going well." I responded, "Do you really know that things aren't going well, and is the something that you are going to do likely to actually improve the situation?" Performance measurement is an important part of investment management and fiduciaries do have a responsibility to assess the progress of the funds for which they are responsible. But, in doing so, they need to carefully understand the limitations of the numbers and ask my two questions of themselves whenever contemplating changes. As is often the case with investing, all of the data in the world and sophisticated tools do not provide the answers which leaves us in the unsatisfying position of making highly subjective decisions.

DIVERSIFIED TRUST

ATLANTA

400 Galleria Parkway, Suite 1400
Atlanta, GA 30339
Phone: 770.226.5333

GREENSBORO

701 Green Valley Road, Suite 300
Greensboro, NC 27408
Phone: 336.217.0151

MEMPHIS

6075 Poplar Avenue, Suite 900
Memphis, TN 38119
Phone: 901.761.7979

NASHVILLE

3102 West End Avenue, Suite 600
Nashville, TN 37203
Phone: 615.386.7302