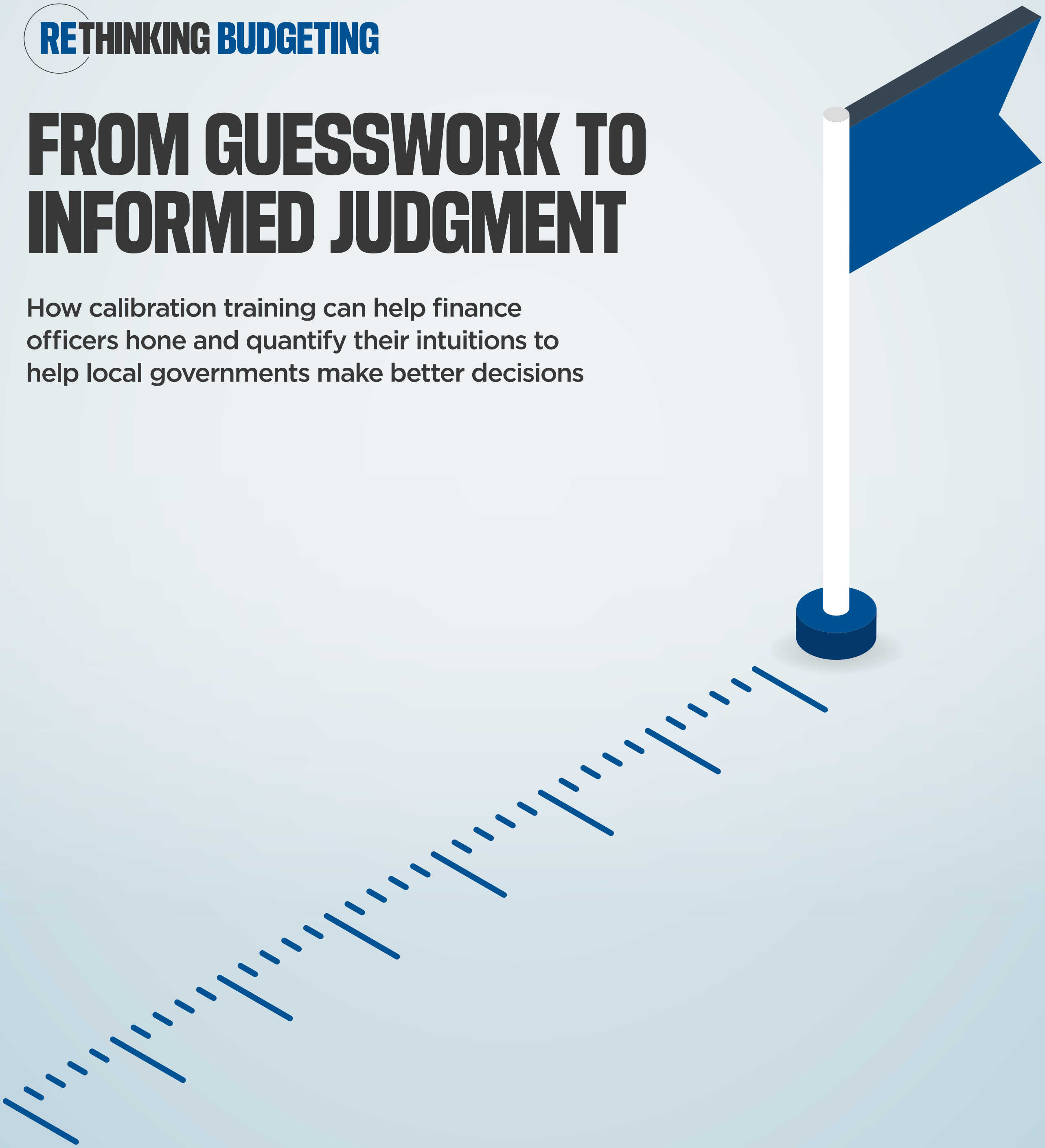
RETHINKING BUDGETING

FROM GUESSWORK TO INFORMED JUDGMENT

How calibration training can help finance officers hone and quantify their intuitions to help local governments make better decisions



ABOUT THE AUTHORS

- **Shayne C. Kavanagh,** Senior Manager of Research, Government Finance Officers Association (GFOA)
- **Doug Hubbard**, CEO, Hubbard Decision Research
- **Philip Martin**, COO, Hubbard Decision Research
- **Robert Weant**, Sr. Analyst, Hubbard Decision Research

ACKNOWLEDGMENTS

We'd like to acknowledge the following people for their review and commentary on the paper:

- **Ariane Gibson**, VP of Financial Services, Utah Housing Corporation
- **Kevin Bueso**, CFO, Regional Transit Agenda, IL
- **Angele Ozoemelam**, Finance Director, City of Tucson, AZ
- **Jim Seaman**, Director of Finance/Assistant Borough Manager, Borough of Media, PA
- **Brian Wilson**, Treasurer/CFO, Hilliard City School District, OH
- **Harpreet Hora**, Executive Director - Budget Services, Atlanta Public Schools, GA
- **Steffanie Dorn**, Treasurer, Greenwood County, SC

ABOUT GFOA

The Government Finance Officers Association (GFOA) represents over 21,000 public finance officers throughout the United States and Canada. GFOA's mission is to advance excellence in government finance. GFOA views its role as a resource, educator, facilitator, and advocate for both its members and the governments they serve and provides best practice guidance, leadership, professional development, resources and tools, networking opportunities, award programs, and advisory services.

ABOUT THE RETHINKING BUDGETING PROJECT

Local governments have long relied on incremental, line item budgeting where last year's budget becomes next year's budget with changes around the margin. Though this form of budgeting has its advantages and can be useful under circumstances of stability, it also has important disadvantages. The primary disadvantage is that it causes local governments to be slow to adapt to changing conditions. The premise of the "Rethinking Budgeting" initiative is that the public finance profession has an opportunity to update local government budgeting practices to take advantage of new ways of thinking, new technologies, and to better meet the changing needs of communities. The Rethinking Budgeting initiative will raise new and interesting ideas like those featured in this paper and will produce guidance for state and local policy makers on how to local government budget systems can be adapted to today's needs. We hope the ideas presented in this paper will spur conversation about the possibilities for rethinking budgeting. The Rethinking Budgeting initiative is a collaborative effort between the Government Finance Officers Association (GFOA) and International City/County Management Association (ICMA).

To learn more, visit gfoa.org/rethinking-budgeting.

The job of a local government finance officer is to help elected officials and other decision-makers make better decisions. This involves helping elected officials make sense of and navigate uncertain information and situations. This requires expressing uncertainty clearly, especially for assumptions that might underlie financial plans and budgets.

To help, GFOA worked with Hubbard Decision Research (HDR) and 40 GFOA member volunteers to go through what is known as "calibration training." The objective of calibration training is to prepare people to express their estimates of uncertainty ***in quantitative terms, using only their own judgment and <u>without</u> being forced to rely on outside data.*** Here are two illustrations of how a quantified estimate of uncertainty might differ from an unquantified estimate.

Unquantified Estimate	Quantified Estimate
Next year's total revenue will be **somewhere around $20 million.**	There is a 90% chance that next year's revenue will be **between $19 million and $21 million.**
I think **it is likely** that the new sales tax law will pass by the start of our next fiscal year.	I give **an 80% chance** that the new sales tax law will pass by the start of our next fiscal year.

In this paper, we will address why quantified estimates of uncertainty are desirable and why special training is necessary to make such estimates. We'll then provide an overview of how the training works. In the first two sections, we will show results from our test group of GFOA members who took the training. In the final section, we will review some of the practical applications of calibration training.

THE FINANCE OFFICER AS A DECISION ARCHITECT

You can read GFOA's **"Budget Officer as Decision Architect"** for a complete discussion of the finance officer's role in helping their governments make better decisions. The ideas in this paper support the skills needed to be a decision architect.

Why Do We Need to Quantify Uncertainty and Training to Do So?

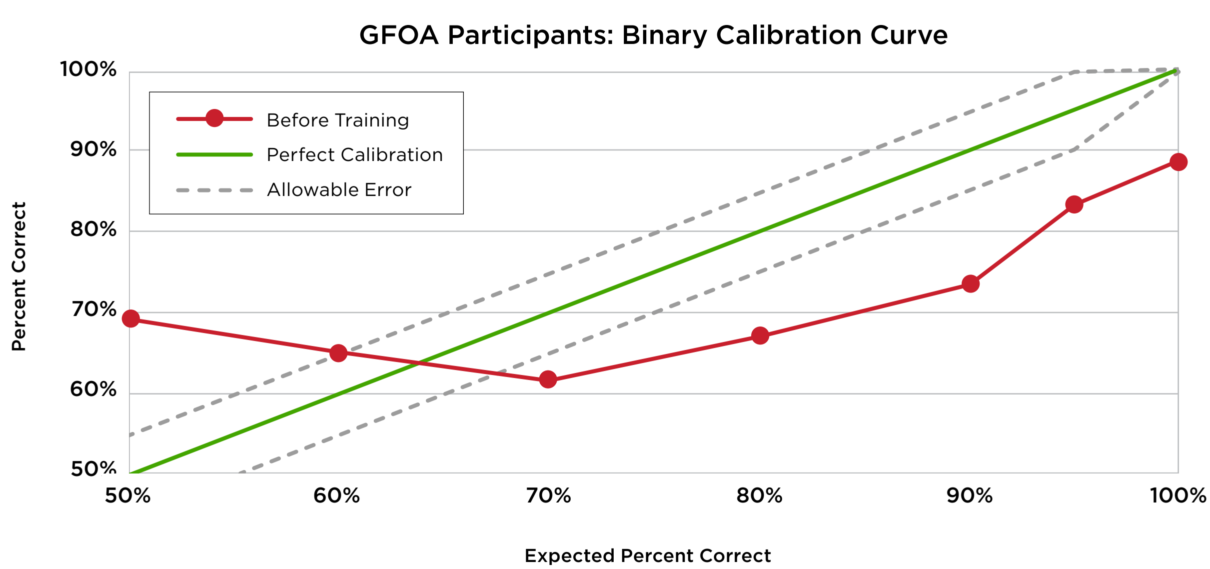
Let's start with why quantified estimates of uncertainty are desirable. Uncertainty is a key element of risk. Risk could be defined as the downside of uncertainty. Finance officers help decision-makers manage the financial risk of their decisions. But, as Peter Bernstein put it in *Against the Gods: The Remarkable Story of Risk*...

> *"Without numbers, there are no odds and no probabilities; without odds and probabilities, the only way to deal with risk is to appeal to the gods and the fates. Without numbers, risk is wholly a matter of gut."*

Bernstein is advocating for quantifying uncertainty (and, thereby, risk) by using odds and probabilities. But why does he caution against our gut instincts? It is because people are not wired to think about risk correctly. One reason is what psychologists call the "overconfidence bias." This means that we are overconfident in our predictions and tend to underestimate uncertainty. We can illustrate this with data from the calibration training experiment we conducted.

Exhibit 1 provides a comparison of how well our group of GFOA volunteers thought they did on making a prediction versus how well they actually did, *before they went through calibration training*. For instance, if we look at the horizontal axis and find the 0.8 mark (or 80%), this is the point where someone believes their forecast has an 80% chance of being correct. We can then look upward until we intersect the 0.8 mark on the vertical axis, which is on the green line. This is the point of perfect calibration: If you say your forecasts have an 80% chance of being correct, then your forecasts are, in fact, correct 80% of the time. Any point on the green line is perfect calibration. We also see a red line on Exhibit 1. This is data from participants before they took the calibration training. If we revisit 0.8 on the horizontal axis, we see the red line meets the vertical axis at 0.68. In other words, the uncalibrated person believed their forecasts had an 80% chance of being correct but are actually correct only 68% of the time. This is overconfidence bias. Anytime the red line is below the green line, it represents overconfidence. The dotted gray lines represent allowable error in our measurement of the participants in the training. We can't expect our tests to be a perfect measure of accuracy.

EXHIBIT 1 | RESULTS OF "PRETEST" OF CALIBRATION TRAINING PARTICIPANTS



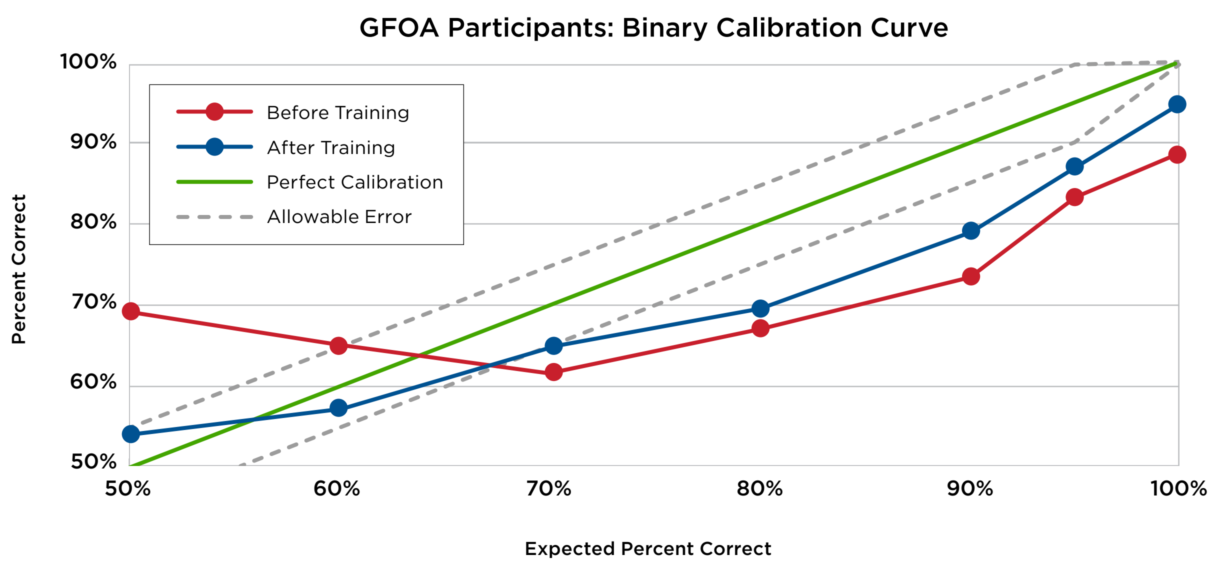
With this in mind, we can see that the participants are not overconfident when assessing their chances at 50% or 60%. This is because the participants are highly uncertain and are, in essence, flipping a coin when estimating if they will be correct or not. In fact, because the red line is above the green line, that means our participants were underconfident: They thought they were flipping a coin; but, in fact, they had better information about the correct answer than they thought they did. But as soon as the participants believe their estimate has a reasonable chance of being right, overconfidence creeps in.

Mitigating or eliminating overconfidence bias is why calibration training is necessary.

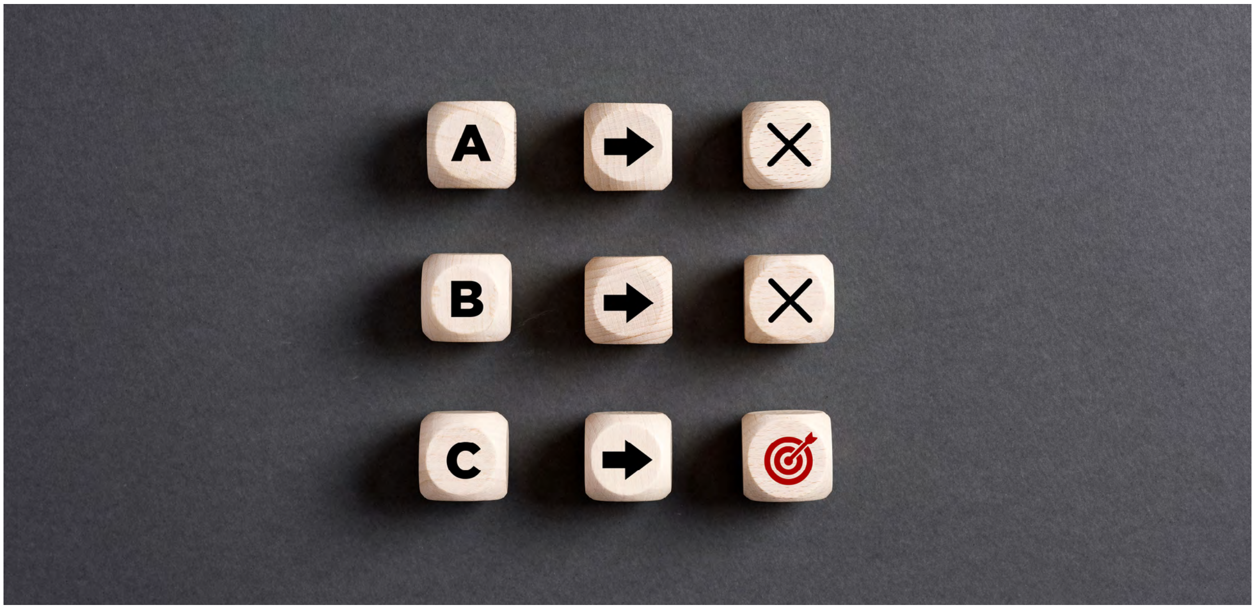
Mitigating or eliminating overconfidence bias is why calibration training is necessary. Quantifying your estimates of uncertainty is not simply a matter of replacing words such as “likely” or “probable” with numbers like “80%” or “4 in 5 chance.” We need to take steps to make the numbers as accurate as possible. Ideally, we would use data to come up with our estimates of uncertainty. For example, if your annual hotel tax forecast has been within +/– 5% of actual revenue for each of the last 10 years, then it is a good bet that next year’s forecast will be within that range as well.* However, the data for estimating uncertainty is often unavailable or incomplete, and there isn’t the time and/or resources to get more data. In this case, the estimator will be forced to rely on their judgment, at least in part.

Calibration training moves the participants closer to the green line. Exhibit 2 shows results from training with an added blue line. We can see that, at all points, the blue line is closer to the green line and is even within the gray lines a few points, which indicates perfect calibration.

EXHIBIT 2 | RESULTS OF “POSTTEST” OF CALIBRATION TRAINING PARTICIPANTS



* Assuming your forecast method remains substantively similar.



How Does Calibration Training Work?

Calibration training helps people recognize overconfidence bias and mitigate it. It does this by providing immediate feedback on the quality of a high volume of predictions made by the participant. In other words, the participants make a lot of predictions and then find out how well they did right away. Rapid feedback is essential to learning.

The predictions that participants make come in the form of difficult "trivia questions," where it is highly unlikely that the participants know the right answer (without looking it up).

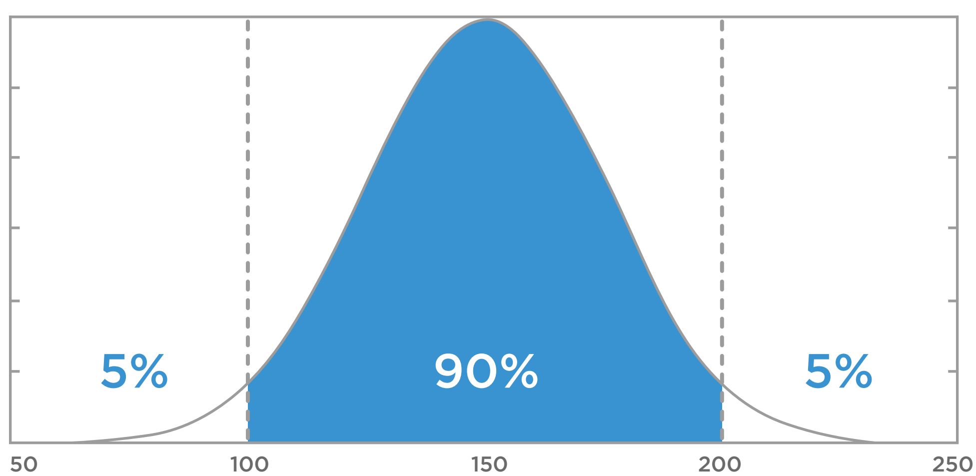
These questions were asked in two formats. One format was true/false, where participants picked either true or false as their answer and then assigned a level of confidence to their answer. For example, a question might be: "True or false? A gallon of oil weighs less than a gallon of water." Though you may not know the exact weight of either, you may recall that oil floats on water; therefore, oil probably weighs less. For that reason, you predict the statement is true and give your answer a high level of confidence (90% or even 100%). Other questions might be harder. For example, "Mars is always farther away from Earth than Venus." Imagine you have no idea, so you guess true...but you rate your confidence at 50%, which indicates you basically flipped a coin to get your answer. Most questions are somewhere in between, where the participant has some idea of the correct answer but is not highly certain. Participants usually overstate their confidence at the start of the training and learn to be more circumspect about their predictions as the training progresses. By answering many questions and getting feedback, participants learn to more accurately assess the chance that they are correct. This prepares them to develop real-life predictions, like whether some important event will come to pass (e.g., whether a major retailer in the community will close in the next year).

The other format of the trivia questions is where participants are asked to define a range where they are 90% sure the correct answer lies within the range. For example, "What is the wingspan, in feet, of a Boeing 737 aircraft (the type of plane used by Southwest Airlines)?" Unless you are an aeronautical engineer, you probably don't know the exact answer and would need to make a projection. You are not making a wild guess because you have information to go on. You may have been a passenger on such a plane or have seen it in pictures, so you have an idea of how big it is. However, it will still be a highly uncertain forecast for you. Picking a single number, like "75 feet," is of limited use because it doesn't express your uncertainty and the degree of risk that you might be wrong. Hence, the calibration training asks participants to express their forecasts as a 90% confidence range. Put another way, they pick a high and a low value where they are 90% confident that the value will be in between. For example, they might say, "I'm 90% confident that the wingspan is between 100 and 65 feet." As it turns out, the actual wingspan is 113 feet. This is a common problem participants have, where they make the range too narrow; this is overconfidence bias in action. When participants first try forecasting 90% confidence ranges for a large number of trivia questions, it is common that the correct answer will fall outside of their ranges for around half of their estimates. If they are truly "90% confident," then the correct answer should fall outside of their ranges in only 10% of their forecasts (90% of the forecasts should be correct). After one half-day of calibration training, participants get much closer to 90% correct for their range questions. This prepares them to develop ranges for situations they might encounter in their work life, such as a range of potential yields for a revenue source.

By answering many questions and getting feedback, participants learn to more accurately assess the chance that they are correct. This prepares them to develop real-life predictions, like whether some important event will come to pass.

Participants do not learn purely through trial and error. They also learn techniques to help them make better predictions. For example, they learn to estimate each side of their ranges separately. When people think of a "90% confidence range," they tend to generate both ends of the range simultaneously. However, considering each side separately causes people to slow down and give more consideration to their estimate (see Exhibit 3). Looking at one side of the range at a time reframes the questions from "90% of the time the correct answer will be in this range" to: A) "Only 5% of the time the correct answer will be higher than my upper value"; and B) "Only 5% of the time the correct answer will be lower than my low value."* Five percent is only 1 out of 20! When participants consider that the correct answer can only go beyond a given side of their interval 1 out of 20 times, it causes them to adjust that side of their interval.

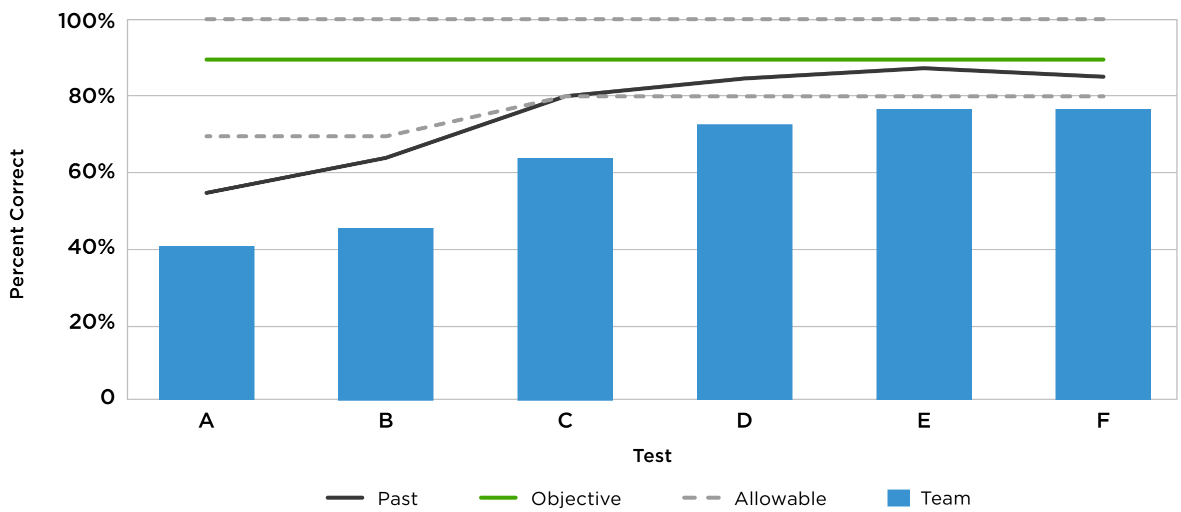
EXHIBIT 3 | 90% CONFIDENCE INTERVAL SHOWN AS A BELL CURVE



* The 5% on either side of the range adds up to 10%. 100% minus 10% is 90% for your 90% confidence interval.

In Exhibit 4, we can see how participants improved as they completed multiple rounds of feedback and incorporated the ideas from the training in their estimates. The graph focuses on the range questions, and the "percent correct" means the proportion of times that the value fell within the 90% confidence interval that the participant defined. By the sixth round, the participants were quite close to the gray line, which is the allowable error. Finally, some readers may wonder why the "objective" in Exhibit 4 (the green line) is only 90% correct instead of 100% correct. This is because scoring 100% on the test is easy; just make all of your ranges infinitely wide. For example, you might set your range for the wingspan of a 737 aircraft between 1 and 1 million feet. You are guaranteed to be "correct." However, such wide ranges do not have much, if any, practical use in real-life estimation problems. Hence, the training aims to create balance, where the ranges are narrow enough to focus in on the 90% most likely possible values but still wide enough to not fall victim to overconfidence bias.

EXHIBIT 4 | PERCENT QUESTIONS CORRECT INCREASES OVER SUCCESSIVE PRACTICE ROUNDS



Let's move into reviewing three practical applications of calibration training.

Applications of Calibration Training to the Role of the Finance Officer

In this section, we'll discuss four possible applications of calibration training for the work of the finance officer. We invite you to also consider other applications.

First, it equips finance officers to communicate uncertainty in their forecasts and estimates. Research shows people prefer advisors who quantify their uncertainty but are still confident.[1] To illustrate, a revenue forecast might be presented in these ways:

- "I'm 75% certain that revenues will increase by at least 1% next year."
- "I'm 90% certain that revenues will be between $50 million and $55 million next year."

You will notice that our examples omit hedging language like "maybe," "I'm not sure, but...," etc. The statements come across as confident, even though expressed as a probabilistic likelihood.

We can also use the second bullet to illustrate overconfidence bias. Overconfidence causes people to make their ranges too narrow, often by around 50% too narrow.[2] Let's revisit the second example in our bullet points. Imagine the person who made that estimate is **calibrated**. In contrast, an **overconfident** estimator might say, "I'm 90% certain that revenues will be between $51 million and $54 million." This is a range of only $3 million compared to the $5 million range that was given by the calibrated estimator. Thus, the overconfident estimator has left less room for error and increased the chance of a bad decision. Calibration training helps you reach intervals that are of the right size given your degree of certainty for the question at hand.

For many applications in public finance, the government needs to settle on a single number. For example, you can't put a range of possible revenues in your budget. *However, you can use the range to assess the risk you are taking on by adopting any given number for your budget.* The GFOA book *Informed Decision-Making through Forecasting: A Practitioner's Guide* provides several case examples of how ranges have been applied by local governments when presenting forecasts. To summarize one application, the range of possible future revenues could be presented to the governing board, with the midpoint of the range considered the single most likely outcome. If the board were to adopt an expenditure budget equal to this midpoint, they give themselves 1-1 odds (or a 1 in 2 chance) of a deficit. This is because there is a 50% chance of revenues being less than (or more than) the midpoint of the estimate. Many elected boards would not like those odds and would prefer to give themselves a better chance of avoiding a deficit at the end of the year. In one of our case studies, the finance officer facilitated a conversation about the odds the elected board would prefer. The board landed on a spending plan that provided 2-1 odds (or a 2 in 3 chance) of producing a surplus, which was more aligned with the board's goal of building up reserves after a recent natural disaster. So they landed on a single number for the budget but were also fully aware of the degree of risk that number represented. After a few years, when the reserves had been replenished, the board opted for lower odds of budgetary surplus (though still better than 1 in 2) by picking a higher number for the expenditure budget. You can download a spreadsheet to help you with the process described in this paragraph as part of an online article published by GFOA called **"Silicon Valley Bank and Stress Tests: What Can Local Governments Learn?"** After you look at the spreadsheet, you will probably be able to think of other applications for inserting subjective estimates of probabilities into larger quantitative models.

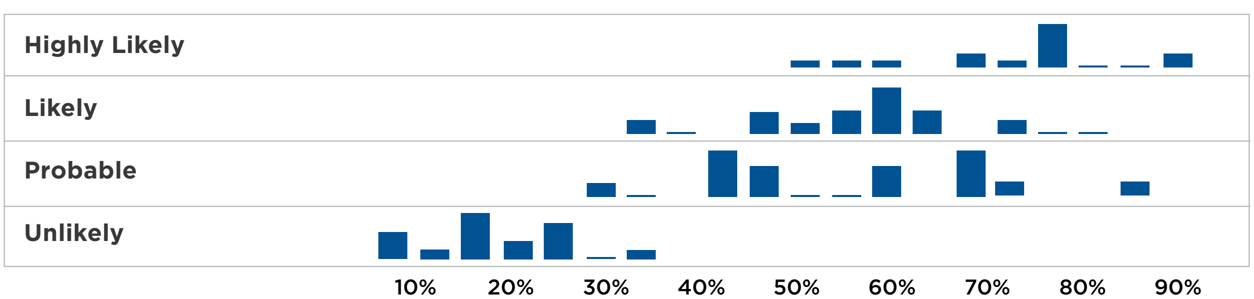
> The illusion of communication is when people use vague terms to communicate uncertainty, such as "likely," "probable," etc. The problem is that people often have very different numbers in mind when asked to quantify these terms.

PRO TIP FOR COMMUNICATING UNCERTAINTIES

Probabilities, like "90%," are an abstract concept. As such, it can be difficult to grasp for people not accustomed to thinking in probabilities. Research suggests that ratios are easier for people to grasp.[3] For example, rather than saying, "I'm 90% certain revenues will be…," one could say, "There is a 9 in 10 chance revenues will be…."

The second practical application of calibration training is that it helps the finance officer avoid the "illusion of communication" when discussing risks. The illusion of communication is when people use vague terms to communicate uncertainty, such as "likely," "probable," etc. The problem is that people often have very different numbers in mind when asked to quantify these terms. To illustrate, one study of military intelligence examined a set of standard terms that intelligence officers were supposed to use to describe their certainty of military threats.[4] The order from most to least certain was: 1) "highly likely," 2) "likely," 3) "probable," and 4) "unlikely." When the intelligence officers were asked to quantify their uncertainty, the results exposed these standardized terms as worthless. For example, it turned out that some intelligence officers thought "highly likely" meant anything with over a 50% chance of occurring, while others thought it meant a 90%+ chance. "Likely" and "Probable" meant essentially the same thing, given the quantities the officers assigned. And the range variation in the quantitative definitions the officers came up with resulted in some officers using "Probable" (the term intended to describe the second lowest chance) to describe the same percent chance that other officers described as "highly likely" (the term for the highest chance). You can see this illustrated in Exhibit 5.

EXHIBIT 5 | FREQUENCY WITH WHICH MILITARY INTELLIGENCE OFFICERS ASSIGNED QUANTITATIVE VALUES TO SUBJECTIVE DESCRIPTIONS OF RISKS



The blue bars represent how often a given quantitative measure (e.g., 20%) was believed to describe what a given subjective term (e.g., "Unlikely" meant). The higher the bar, the more frequent the belief.

The illusion of communication is not limited to military intelligence. Many local governments have probably fallen into the same trap by using the popular "heat map" style of risk matrix, as shown in Exhibit 6. It is easy to imagine the same problem with the terms on either axis of Exhibit 6. This is not just a theoretical problem. Research into how these kinds of risk matrices impact real-life decisions has concluded that "[qualitative] risk matrices should not be used for decisions of any consequence."[5] Calibration training prepares the finance officer to replace vague terms with meaningful, quantified, and calibrated estimates of uncertainty. For example, you could imagine a version of Exhibit 6 where the two axes are replaced by ranges of precent likelihood that the risk will occur and a potential dollar loss on the other. There are also other presentation methods that quantify risks into probabilities as a fundamental design feature. One example is called a "loss exceedance curve." You can see examples applied to cyber security risk in the GFOA report Cyber Risk Savvy.

The third practical application of calibration training is that it hones our intuitions about chance and prepares us to think probabilistically. According to The Great Mental Models project:[6] "Successfully thinking in shades of probability means roughly identifying what matters, coming up with a sense of the odds, doing a check on our assumptions, and then making a decision. We can act with a higher level of certainty in complex, unpredictable situations. We can never know the future with exact precision. Probabilistic thinking is an extremely useful tool to evaluate how the world will most likely look so that we can effectively strategize."

The benefits of probabilistic thinking are measurable. For example, one study showed that exposing participants to probabilistic training was associated with as much as a 50% increase in the accuracy of their predictions.[7] Having a general idea how a revenue might perform is one thing. However, being able to examine the question of revenue yield from multiple vantage points and think about how new information might cause you to adjust your probabilistic expectations can be much more powerful.

EXHIBIT 6 | A CONVENTIONAL RISK MATRIX

RISK PROBABILITY	RISK SEVERITY				
	Catastrophic A	Hazardous B	Major C	Minor D	Negligible E
Frequent 5	5A	5B	5C	5D	5E
Occasional 4	4A	4B	4C	4D	4E
Remote 3	3A	3B	3C	3D	3E
Improbable 2	2A	2B	2C	2D	2E
Extremely improbable 1	1A	1B	1C	1D	1E

Conclusion

Public finance officers must deal with uncertainty and risk across many domains of public finance, including forecasts, rainy day funds, insurance, and more. The ability to think probabilistically enhances a finance officer's ability to make savvier decisions about risk and uncertainty. Calibration training hones probabilistic thinking skills. Even better, it provides a basis for communicating quantified probability estimates to others. These skills are essential for helping the finance officer guide their local government in testing assumptions in an uncertain world. This allows the finance officer to be a bona fide decision leader—someone who doesn't only make good decisions themselves but also can help others make wise decisions.[8]

ENDNOTES

[1] Gaertig, C., & Simmons, J.P. (2017). Do people inherently dislike uncertain advice? Forthcoming, Psychological Science. Available at SSRN: https://ssrn.com/abstract=3041566

[2] Soll, J., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 299–314.

[3] Heath, C., & Starr, K. (2022). *Making numbers count: The art and science of communicating numbers*. Avid Reader Press / Simon & Schuster.

[4] Heuer, R. (1999). *The Psychology of Intelligence Analysis*. Center for the Study of Intelligence, CIA.

[5] Thomas, P., Bratvold, R. B., & Bickel, J. E. (2013). The Risk of Using Risk Matrices. SPE Economics and Management. 6(02)

[6] *The Value of Probabilistic Thinking: Spies, Crime, and Lightning Strikes*. Farnam Street Media Inc. https://fs.blog/probabilistic-thinking/#:~:text=It%20is%20one%20of%20the,be%20more%20precise%20and%20effective.

[7] In *Smarter Faster Better: The Transformative Power of Real Productivity* (Random House Publishing, 2016), Charles Duhigg discusses the results obtained by the “Good Judgment Project” in their forecast experiments.

[8] Decision leader is a term from: Moore, D. A. & Bazerman, M. H. (2022). *Decision leadership: Empowering others to make better choices* (11–12). Yale University Press. Kindle Edition.

RETHINKING BUDGETING

gfoa.org/rethinking-budgeting

Government Finance Officers Association
203 N. LaSalle Street, Suite 2700
Chicago, IL 60601
312-977-9700 | gfoa.org