

## Stop Using Relative Values THEY DON'T WORK AS ADVERTISED

Ryan T. Baker



## Abstract

Marine Corps selection boards rely heavily on fitness report relative values (RVs) when considering Marines for promotion or special assignment, but RVs are poorly understood metrics that do not work as advertised. They are, in fact, quirky numbers that misrepresent the location of fitness reports in reporting senior (RS) profiles in ways that selection boards cannot currently detect. This problem has gone largely unnoticed because very few people—not even members of selection boards—are allowed to see both the RVs and the RS profiles they summarize. In this paper, I show RVs alongside the corresponding RS profiles to demonstrate just how misleading they can be. I also draw on ideas from the field of statistics to show that RVs are ill-suited to their task, and that in small RS profiles, they are no more reliable than random statistical noise. I also provide several recommendations for addressing these problems—the most important of which is to stop using RVs altogether.

## Acknowledgments

The author would like to thank Dr. Kim Deal, Mr. Jonathan Geithner, Dr. Jennifer Griffin, Colonel Jeffery M. Peterson (USMC, Ret.), Colonel Dave Winnacker (USMC), Lieutenant Colonel Joe Mason (USMC), Lieutenant Colonel Ryan Pallas (USMC), Major Katy Evezich (USMC), Major Bob Jankowski (USMC), Major Wing Kwan (USMC), Major Megan MacDonald (USMC), and the Manpower and Defense Workforce team at CNA for their help, encouragement, and critical feedback. Any remaining errors are the author's alone.

## Disclaimer

The author is both a CNA research scientist and a reserve member of the US Marine Corps. This paper reflects the author's personal views and does not necessarily reflect the views of the Marine Corps or the Department of Defense. For transparency, we note that the author has occasionally directly participated in the process he critiques in this paper—as an author and recipient of fitness reports, and as a member of a selection board.

**DISTRIBUTION STATEMENT A.** Approved for public release: distribution unlimited.

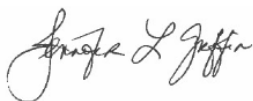
Public release

1/12/2024

**Cover image:** US Marine Corps photograph by Sgt Erin Morejon.

This document may contain materials protected by the Fair Use guidelines of Section 107 of the Copyright Act, for research purposes only. Any such content is copyrighted and not owned by CNA. All rights and credits go directly to content's rightful owner.

**Approved by:**



**January 2024**

Jennifer Griffin  
Research Program Director  
Resources and Force Readiness Division

Request additional copies of this document through [inquiries@cna.org](mailto:inquiries@cna.org).

## TABLE OF CONTENTS

Introduction.....	1
What Is an RV? .....	3
RVs Are Often Ambiguous and Misleading.....	5
RVs Make Assumptions That Boards Can't Verify .....	8
Assumption 1: A midrange profile average ...	8
Assumption 2: Meaningful range.....	10
Assumption 3: Large sample size.....	12
RVs Are Too Important to Ignore .....	15
RVs Make Prohibited Comparisons .....	17
A Graphical Approach Would Be Better.....	19
Concluding Remarks.....	23
Appendix A.....	24
Appendix B.....	25

# INTRODUCTION

Fitness report relative values (RVs) are a peculiar feature of the Marine Corps' performance evaluation system (PES). Few understand how they are calculated, but they play an important role in the way the Marine Corps selects Marines for promotion and special assignments.

However, these values are unreliable. RVs mask important information from selection boards and misrepresent the location of reports within reporting senior (RS) profiles. In other words, they don't work as advertised, and the Marine Corps should stop using them.

This problem has gone largely unnoticed because the structure of the PES makes it hard to see. Right now, almost no one is allowed to see RVs alongside the RS profiles they summarize. RSs can see their profiles, but not the RVs of the reports they write; conversely, selection boards can see the RVs, but not the RS profiles they were derived from. If anyone were to put the two together, they would see that RVs are quirky numbers that obscure important information. They would see, for example, that a report can simultaneously be in the top third of the RS profile and in the bottom third of the RV scale, that the middle "third" can be empty, and that a report's cumulative RV can increase even as it remains the lowest report in the profile. Just as important, they would realize that the current process makes it



RVs mask important information from selection boards and misrepresent the location of reports within reporting senior (RS) profiles.

impossible for selection boards to detect these quirks from inside a board room.

Correcting this problem has a new urgency given the importance of successful talent management to the Marine Corps' future operating concepts. As the 38th Commandant put it, "Our modern operational concepts and organizations cannot reach their full warfighting potential without a talent management system that recruits, develops, and retains the right Marines."<sup>1</sup> Unfortunately, the only way to retain the right Marines in the current "up or out" system is for selection boards to identify and select them. RVs make that task harder.

In 2015, I coauthored a short article for the *Marine Corps Gazette* that discussed some problems with RVs and proposed ways to address them, but that article failed to prompt much change and only scratched the surface of the problem.<sup>2</sup> Since RVs

<sup>1</sup> David H. Berger, *Talent Management 2030*, Nov. 3, 2021, 3. Later, the Commandant argues, "Today we are out of balance, placing too much emphasis on recruiting new personnel to maintain end strength, and too little emphasis on identifying and retaining the most talented individuals already in our ranks" (p. 6).

<sup>2</sup> Maj Ryan T. Baker and Capt Yuk Wing Kwan, "Miscalculating Performance: How Relative Values Work, and Why the Marine Corps Should Stop Using Them," *Marine Corps Gazette* 99, no. 12 (Dec. 2015): 64–68. None of the recommendations we proposed in 2015 were implemented. Instead, the head of the Performance Evaluation Section at Manpower and Reserve Affairs (MMRP-30) wrote a separate article arguing that our recommendations were unnecessary. See LtCol Christopher D. Pritchett, "Rebuttal to 'Miscalculating Performance': Better Understanding the Complexities of the System," *Marine Corps Gazette* 100, no. 11 (Nov. 2016): 62–63. See also LtCol Pritchett's letter to the editor responding to two other articles (and expressing regret for publicly disclosing the formula for relative values) in *Marine Corps Gazette* 101, no. 6 (June 2017): 6.

remain poorly understood and their shortcomings still largely unacknowledged, a more detailed follow-up is necessary—hence this paper.

I am not going to retread all the ground covered in the 2015 article, but I will review some of the key ideas on the way to presenting several new ones. My approach is straightforward: I show RVs alongside the RS profiles they were derived from.<sup>3</sup> In the tables and figures that follow, I present examples that show what selection boards currently cannot see, revealing just how ambiguous and misleading RVs can be. Later in the paper, I draw on ideas from the field of statistics to show that RVs in small profiles are often statistically indistinguishable from random noise, and that the way selection boards use RVs involves a well-known issue with data summaries that statisticians have warned people about for decades. I will also highlight an important tension between RVs and the legal requirements for promotion boards.

Based on my analysis, I make three major recommendations:

- 1. Stop using relative values.** RVs should not be used by selection boards or as a general-purpose measure of performance. There are better and more transparent ways to weigh fitness reports.
- 2. Display RS profiles graphically on the master brief sheet (MBS).** Many RS profiles are too small for summary statistics to be useful, and report averages are often distributed in ways that selection boards should know about when weighing fitness reports. I provide an example of a graphical MBS later on.

- 3. Separate active and reserve RS profiles.** The Marine Corps is required to consider active and reserve Marines separately for promotion, but RVs and RS profile dynamics force active and reserve Marines into competition with one another in the board room.

To be clear, this paper focuses on only RVs—not on other issues with the fitness report system. For example, I do not address the tension between the attribute mark descriptions and the logic of an RS marking philosophy; the contradiction between the distributional assumptions of RVs and the comparative assessment; or how boards should weigh reports from excessively short observation periods, inconsequential billets, periods of academic instruction, or similar occasions. In this paper, I focus narrowly on the problems that arise from RVs and how best to address them, with the understanding that other issues also merit attention.

For brevity, I will assume the reader is familiar with the PES and the Marine Corps' selection board process. For those unfamiliar, I recommend reviewing the relevant order (MCO 1610.7B, commonly referred to as the "PES Manual") and the educational materials on the websites of the Performance Evaluation System Unit (MMPB-23) and Officer Promotions Unit (MMPB-10). Although few of these references are free of errors (at least not as of this writing), they contain enough background to guide the reader.<sup>4</sup>

<sup>3</sup> The Performance Evaluation Section of Manpower Management Support Branch (MMSB-30, since reorganized into MMRP-30, and then again into MMPB-23) published the equation for RVs in 2013, although Manpower has since removed it from their publicly available material. See MMSB, "PES Brief for MROs and Reporting Officials," (PowerPoint presentation, Apr. 12, 2013), slide 26. For the full conditional function, see Baker and Kwan, "Miscalculating Performance," 67. See also Adam Clemens, Lauren Malone, Shannon Phillips, and Gary Lee, *An Evaluation of the Fitness Report System for Marine Officers*, DRM-2012-U-001003-Final, CNA, 2012, 8n5.

<sup>4</sup> See US Marine Corps, *Performance Evaluation System (PES)*, MCO 1610.7B, June 2023 (hereafter cited as "PES Manual"). The online materials are undated; the versions cited here were downloaded in September 2023.

# WHAT IS AN RV?

RVs are derived from the attribute marks on fitness reports.<sup>5</sup> An observed report can have up to 14 of these marks, which are averaged together to produce a fitness report average (FRA, or “RPT Avg” on an MBS). Any attributes marked “N/A” are excluded from the calculation.

Each RV is a number between 80 and 100 that shows how the FRA on a given report compares to two other numbers: (1) the average marks an RS has given to Marines of the same grade, and (2) the highest marks the RS has given to a Marine of the same grade.<sup>6</sup> An RV above 90 indicates a report’s attribute marks are above average for that RS; an RV below 90 indicates the marks are below average. A report with an RV of 100 has the highest marks in the profile (or is tied for the highest).<sup>7</sup>

RVs appear in only one place—the MBS of the “Marine Reported On” (MRO)—and come in two varieties: the RV at processing and the cumulative RV. The RV at processing is a fixed value that will never change. It shows the RV when the report was first processed by Headquarters, Marine Corps (HQMC). The cumulative RV is a dynamic value that can change every time a new report is added to the



## WHAT IS AN RV?

RVs are derived from the attribute marks on fitness reports.<sup>5</sup> An observed report can have up to 14 of these marks, which are averaged together to produce a fitness report average.

RS profile. The RV at processing and the cumulative RV can be the same, but they need not be.

Individual RVs are printed on the MBS alongside other information derived from the fitness report (see Figure 1 for an example), but RVs are also tabulated in a summary table at the top of the MBS. This table separates reports into “thirds”—upper, middle, or lower—based on where they fall on the scale between 80 and 100 (reports without an RV are counted in an “N/A” column).<sup>8</sup> Selection boards use this table to identify trends in performance over

<sup>5</sup> “Attribute marks” is a convenient shorthand. Technically, each of the 14 attributes is evaluated using Performance-Anchored Rating Scales (PARS). See PES Manual, p. 4-20.

<sup>6</sup> Several notes here: (1) unless indicated otherwise, I use the word *average* to refer to the mean; (2) when all the reports in a profile have the same average, the system assigns every report an RV of 90; and (3) “Marines of the same grade” does not necessarily mean “other Marines of the same grade”—an RS could write three fitness reports on the same Marine, and the system would calculate RVs for those reports.

<sup>7</sup> Once an RS has written at least three observed reports on Marines of the same grade—excluding end-of-service, extended, and academic reports—the system will calculate an RV for each report in that profile. For a more detailed walkthrough of the RV calculation, see the appendix in Baker and Kwan, “Miscalculating Performance.” On the exception for end-of-service and academic reports, see MARADMIN 412/20 and PES Manual, pp. 3-5 and 6-3.

<sup>8</sup> See PES Manual, p. E-1; MMRP, “Master Brief Sheet Overview,” (PowerPoint presentation, Dec. 2022), slides 21–24, [https://www.manpower.usmc.mil/webcenter/portal/MMRP\\_OP](https://www.manpower.usmc.mil/webcenter/portal/MMRP_OP).

## Stop Using Relative Values THEY DON'T WORK AS ADVERTISED

a Marine's career—the higher the proportion of reports in the “upper” column, the better the Marine's performance will appear to a selection board.

RVs, then, are essentially “a time saving metric that prevents board members from having to pull the profile of every RS of every report for

every Marine being considered by the board.”<sup>9</sup> Their “single purpose,” according to HQMC, is “to provide a promotion/selection board with a quick interpretation of a FITREP's placement within an RS's profile.”<sup>10</sup>

---

<sup>9</sup> MMRP, “Performance Evaluation System Part II: New Reporting Officials” (undated PowerPoint presentation), slide 6, <https://www.manpower.usmc.mil/webcenter/portal/MMRP30>. This brief is the second of four educational briefs available on MMPB's website. Hereafter, I will cite these briefs as “PES Brief Part [I-IV]” and provide the part and slide number(s).

<sup>10</sup> PES Brief Part III, slide 10. NB: Descriptions of the purpose of RVs vary across official sources. The PES Manual says the only purpose of RVs is “to give individuals making personnel management decisions the ability to weigh the merit of a single fitness report in relation to the RS's rating history or ‘profile’” (PES Manual, p. 8-6). A third view—one that can be found in old student handouts from The Basic School—is that RVs are designed to “account for difference in grading philosophies” and “control for the variable toughness of reporting seniors.” See “Fitness Reports B3K0477XQ-DM Student Handout,” Basic Officer Course, Dec. 2015, 6.

# RVS ARE OFTEN AMBIGUOUS AND MISLEADING

Unfortunately, RVs are not a reliable guide to the location of a fitness report within an RS profile. The highest marks in the profile will always receive an RV of 100, but in all other cases, RVs are accurate only when several key assumptions hold. Those assumptions (discussed in more detail later) are both easy to violate and impossible for boards to check.

The easiest way to see the unreliability of RVs is with an example. Imagine you are a member of a

selection board trying to make sense of the fitness report in Figure 1. Also imagine this report covers time in combat or a stint in battalion command, so you are especially interested in weighing this report accurately. Let's assume the narrative comments from the RS are lukewarm—full of stale phrases and faint praise, but nothing that would make the report adverse. How should you interpret the report?

Figure 1

REPORTING SENIOR MARKINGS															
Reporting Senior		Per	Pro	Cou	Eff	Ini	Lea	Dev	Set	Ens	Co	PME	Dec	Jud	Eval
Promote	Reports	RPT Avg		RS Avg		RS High		RPT at High			RV at Proc			Cum RV	
		C	C	B	B	C	B	B	B	B	C	C	B	C	C
Yes	3 of 6	2.50		2.99		5.00		1			80			87.56	

Note: An extract from an MBS showing the attribute marks on a fitness report and how they compare to others in the RS profile. The RV is supposed to show the location of the report within the underlying RS profile, but it can't do so reliably (as Table 1 demonstrates).

You can see the report has never been at the top of the RS profile because it had an RV of 80 when it was processed. But you can also see that the RS has written three additional reports that, together, increased the cumulative RV from 80 to 87.56. The report is still below average, but it has moved from the bottom to the middle third. It looks like the initial assessment was artificially low—the RS profile was very small when the report was processed, but the RS pulled this Marine up closer to the middle of the

profile with subsequent reports. Considering the narrative comments and the cumulative RV together, this looks like a “so-so” report; it's not great, but not terrible—and not the worst the RS has written.

This seems like a reasonable assessment. It might even be correct, since the numbers in Figure 1 are consistent with a profile in which an RS has added several lower ranked reports over time. But—it's also consistent with a profile in which the report is (and



has always been) the single lowest ranked report in the profile. Far from trying to pull this Marine up from the bottom, the RS may be trying to push them *further down* as they write additional reports. But because of the way RVs work, the board will be steered toward the first interpretation and away from the second. They will see an evaluation that improved over time, even if the opposite is true.

Table 1 illustrates this possibility. It shows two RS profiles, either of which could have produced Figure 1. In Profile 2, the report with an average of 2.50 is at the very bottom of the RS profile; in Profile 1, it is the *fourth* from the bottom—meaning it's closer to the top than the bottom, by rank.

**Table 1**

	Profile 1	Profile 2
<b>RS Profile</b>		
Report 1	4.00	2.57
Report 2	5.00	2.64
Report 3	<b>2.50</b>	<b>2.50</b>
Report 4	2.08	2.62
Report 5	2.00	2.62
Report 6	2.36	5.00
<b>Master Brief Sheet</b>		
Reports	6	6
RPT Avg	2.50	2.50
RS Avg	2.99	2.99
RS High	5.00	5.00
RPT at High	1	1
RV at Proc	80.00	80.00
Cum RV	87.56	87.56

Note: Two RS profiles, either of which could have produced the numbers in Figure 1. In Profile 1, the report highlighted in pink is above the median. In Profile 2, the same report is last.

This quirk is possible because of the way several features of the PES interact, but two stand out. First, contrary to common belief (and contrary to the PES Manual before 2015), **the lowest ranked report in an RS profile can have an RV higher than 80.** The equation that converts FRAs to RVs keeps track of the highest report in the profile and ensures it receives an RV of 100, but it does *not* keep track of the lowest report and does *not* ensure it receives an RV of 80.<sup>11</sup>

Second, **the RV scale is always symmetric**—the distance from 100 to 90 will always be the same as the distance between 90 and 80. So when the highest report in the profile is replaced by an even higher report (e.g., when you add a “water walker” to your profile), the RV scale extends in *both* directions, which has the counterintuitive effect of raising the RVs of the reports at the bottom of the profile.<sup>12</sup> This example has several important implications that are worth emphasizing:

- Adding a “water walker” to an RS profile will lower the RVs of reports toward the top of the profile and raise the RVs of those toward the bottom. This means it will hurt those who have performed relatively well and help those who have performed relatively poorly.
- In most cases, selection boards cannot tell whether an RV changed because reports were added above or below it in the profile. They cannot see the underlying distribution of reports, only the fitness report itself and the profile information summarized on the MBS.
- It is impossible for RSs to provide narrative comments that clear up this kind of ambiguity. The problem arises over time as new reports are added to a profile, but the

<sup>11</sup> The *RS Summary* report available on Marine Online (MOL) lists the lowest value in the profile, but this information is not used in the calculation of RVs and is not available to selection boards.

<sup>12</sup> “Water walker” is a colloquialism for an outstanding Marine.

narrative comments for each report are fixed when that particular report is submitted.

- Order matters. The RV at processing and the cumulative RV are different only because of the order in which the reports were processed—which means changing the order can often change the way a report appears on an MBS. For example, in Table 1, if the reports in Profile 1 were completed in the reverse order, the RV at processing would have been 100 rather than 80.<sup>13</sup>

- RV “thirds” do not contain an equal number of reports. Unless all the reports in a profile have the same FRA, the math ensures there will always be at least one report at 100 and at least one below 90, but the lowest report can be in the middle third and very close to 90.

The key point is that RVs are ambiguous—the standard, intuitive interpretation may be correct, but the numbers on an MBS will usually be consistent with more than one RS profile, and boards can't know for sure their intuition is correct. Unfortunately, the PES is constructed in a way that makes this situation likely. RVs are only unambiguous when certain assumptions hold, and those assumptions are not easy to meet.<sup>14</sup>

<sup>13</sup> The report in Profile 2 would've moved only slightly, from 86.24 to 87.56. Note that processing order affects only the RV at processing, not the cumulative RV.

<sup>14</sup> An exception: RVs are unambiguous for the highest report(s) in the profile, which will always be 100.

# RVS MAKE ASSUMPTIONS THAT BOARDS CAN'T VERIFY

The way selection boards use RVs involves at least three important assumptions about the underlying RS profile. The first is that the profile average is centered within the profile range (an average that is “midrange”), the second is that a meaningful difference in performance exists between the highest and lowest rated Marines in the profile, and the third is that the profile is a reliable proxy for the RS’s marking philosophy (i.e., a large sample size). When one or more of these assumptions do not hold, RVs will create a distorted picture of relative performance.

## Assumption 1: A midrange profile average

At the top of most MBSs is a summary table that separates fitness reports into thirds using RVs. According to the PES Manual, “A relative value between 93.34 and 100.00 indicates the report is in the upper third of the RS profile. A relative value between 86.67 and 93.33 indicates the report is in the middle third of the RS profile, [and] a relative value between 80.00 and 86.66 indicates the report is in the bottom third of the RS profile.”<sup>15</sup> Unfortunately, **the PES Manual is wrong**. An RV of between 93.34 and 100.00 indicates the report is in the upper third of the *RV scale*, not the RS profile.

The RS profile and RV scale are not the same thing. The RV scale is derived from the RS profile, but the equation for RVs requires them to intersect at only two points: the value of the highest report in the



### Three assumptions selection boards make about the underlying RS profile:

1. The profile average is centered within the profile range.
2. A meaningful difference in performance exists between the highest and lowest rated Marines in the profile.
3. The profile is a reliable proxy for the RS’s marking philosophy (a large sample size).

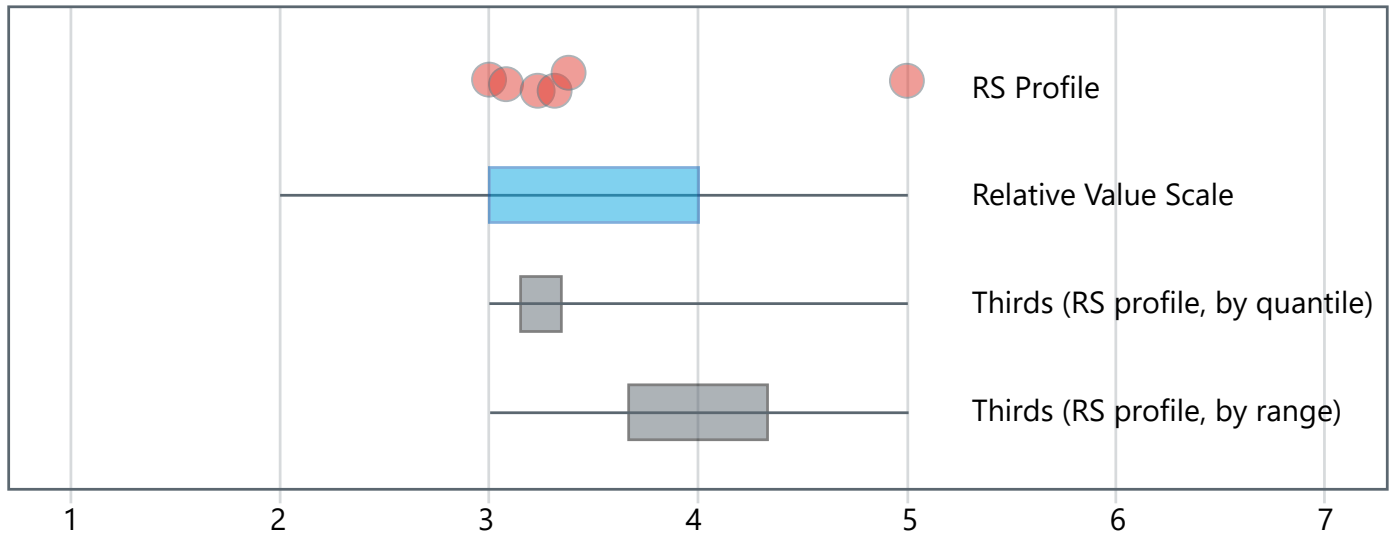
profile and the value of the profile average.<sup>16</sup> That leaves a lot of wiggle room, and there is nothing to prevent a report from being in a different third in the RS profile than on the RV scale.

Figure 2 makes this clear. It compares thirds on the RV scale to thirds within the RS profile. The dots along the top are notional FRAs plotted along a number line (the values are taken from the first column of Table 2). The dots are “jittered” vertically to make the overlapping circles easier to see, but

<sup>15</sup> PES Manual, p. 8-6.

<sup>16</sup> PES Manual, p. 8-6. For the full RV equation, see Baker and Kwan, “Miscalculating Performance,” 67.

Figure 2



Note: A comparison of “thirds” measured three ways. The dots along the top represent a notional RS profile (each dot is one fitness report average). The line immediately below the dots shows the corresponding RV scale with the middle third highlighted with a blue box. Reports to the right of the box are in the top third; reports to the left are in the bottom third. The two lines below the RV scale break the RS profile into thirds directly using two alternative methods (the gray boxes enclose the middle thirds). NB: It’s possible for a report to be in a different third on the RV scale than in the RS profile.

not horizontally—so the averages correspond to the number line at the bottom exactly. The line immediately below the dots shows the corresponding RV scale with the middle third highlighted with a blue box. Reports that fall within the box on the number line are in the middle third, reports to the right of the box are in the top third, and reports to the left are in the bottom third. Note that every report except the highest one is in the middle third on the RV scale. Also note that the RV scale extends well below the range of the RS profile—there are no reports in the bottom third. Contrast the RV scale with the bottom two plots in the same figure, each of which breaks the RS profile into thirds directly without using the RV.

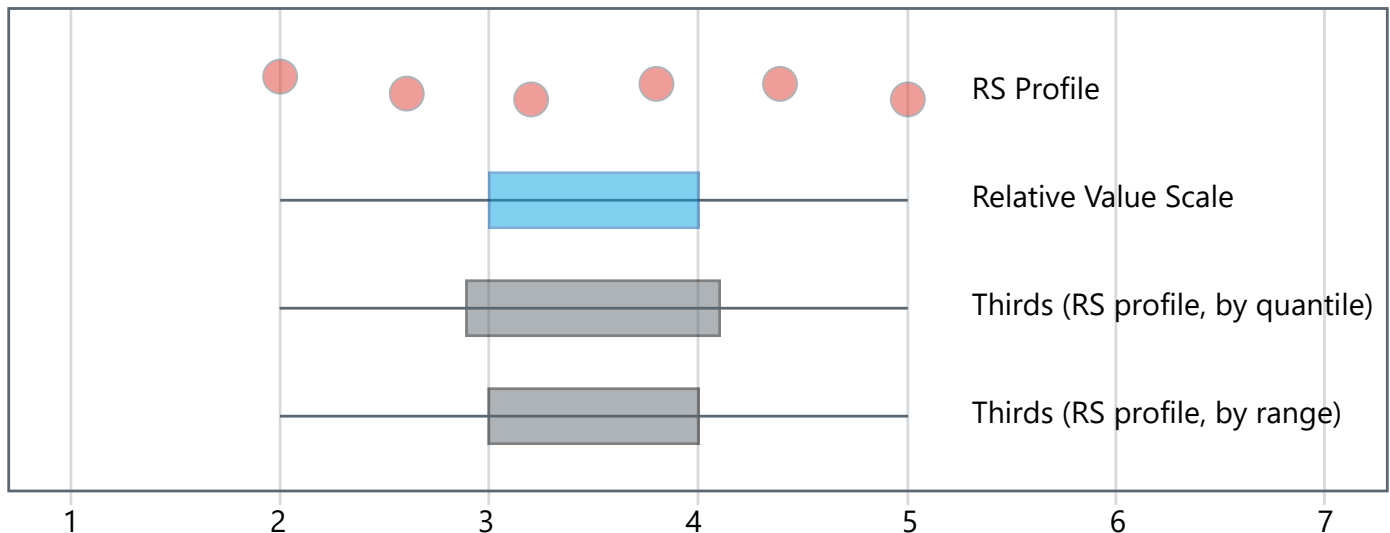
Figure 2 shows two alternative plots because thirds can be reasonably calculated in two ways—based on the number of reports in the profile or based on the range of the reports on the number line. The first ensures the number of reports in each third is equal

(like thirds at The Basic School (TBS));<sup>17</sup> the second ensures each third covers the same distance within the RS profile. The primary advantage of the first approach is ease of interpretation—a report in the middle third will always have reports both above and below it in the RS profile, which is intuitive. The primary advantage of the second approach, using distance on the number line, is that it preserves outlier performance—both good and bad.

In Figure 2, the bottom plot is based on the range of the RS profile, and the one above is based on the number of reports. In both cases, the comparison with the RV scale is stark. Every report except the highest is in the middle third on the RV scale, but every report except the highest is in the *bottom* third of the RS profile (by range). Even in the quantiles plot, fully half the reports are in a different third in the RS profile than on the RV scale, and the second highest report is in a different third in all three plots!

<sup>17</sup> These are called “quantiles” (or, when using exactly three bins, “terciles”).

Figure 3



Note: Compare with Figure 2. We can be sure a report falls in the same third on both the RV scale and in the RS profile only when the distribution of reports is unusually symmetrical.

Despite what the PES Manual says, we can be sure that a report falls in the same third on both the RV scale and RS profile only under certain—very restrictive—conditions. For RV thirds and “by range” thirds to line up, the profile average must be equidistant from the highest and lowest reports in the profile (i.e., at the midrange of the RS profile). For RV thirds and “quantile” thirds to line up, reports must also be evenly distributed throughout the profile. Figure 3 shows what it takes to get all the definitions to yield the same answer.<sup>18</sup>

RS profiles are not always distributed in this way, of course, but selection boards cannot see the profiles to verify one way or the other. Although an RV of 80 is at the bottom of the RV scale, it can be in the middle third of the RS profile with several reports ranked below it. When that happens, the report will be indistinguishable on an MBS from the worst (non-adverse) report an RS has ever written.

Similarly, a middle third report on the RV scale can be the worst report an RS has ever written, and boards would have no way to know. (The single exception is the highest report in the profile, which has an RV of 100 and will always be in the same place in all three calculations.)

## Assumption 2: Meaningful range

A lack of meaningful range occurs when the difference between 80 and 100 on the RV scale does not correspond to a meaningful difference in performance.

Imagine an RS has never written a report on a staff sergeant (SSgt) before but now has two who work for her. For their annual reports, the RS dutifully follows the PES Manual and evaluates the Marines consistent with her personal marking philosophy.<sup>19</sup> In her mind, a “C” is an average mark (which corresponds

<sup>18</sup> Note how different the distribution in Figure 3 looks from the “Christmas tree” distribution in the comparative assessment portion of the fitness report. Ironically, to work as advertised, RVs require a distribution of performance that is substantially different than the one actually printed on the report.

<sup>19</sup> The PES Manual provides a detailed discussion of marking philosophies beginning on page 4-21.

to a 3.00 in fitness report math), and both of her SSGts are slightly above average performers. After completing both reports, the RS checks her profile and verifies that both reports came in just above 3.00, with one slightly higher than the other (3.15 and 3.23, respectively). That makes sense to her. Both SSGts performed above average according to her marking philosophy, but their performance was not identical—one slightly outperformed the other.

Now imagine that a few months later, one of the SSGts is transferred to another section and begins working for a new RS, prompting a transfer (TR) fitness report.<sup>20</sup> Since it's only been a few months and the Marine's performance has not changed much, the RS gives the Marine similar marks on the TR report, improving only a single mark. The Marine's profile now has three reports—one at 3.15, one at 3.23, and one at 3.30—which is enough for the PES to calculate RVs.

This scenario violates the assumption of meaningful range. The report at 3.15—the report intended to capture “above average” performance—will have an RV of 80, the lowest possible value. The report at 3.23 will have an RV of 90, and the TR report at 3.30 will have an RV of 100. In this example, only two steps on the attribute scale account for the entire difference between 80 and 100 on the RV scale. Consequently, the “above average” SSGt with the 3.15 report is stuck with an RV that could only appear worse on an MBS if it were marked adverse. Notably, this outcome will occur regardless of the objective performance of the Marines; the RVs would be the same whether the average attribute mark was a “C” or an “E.”

In this situation, the PES Manual directs RSs to address the issue in the Section I comments:

“Address any conflicts within the fitness report, or the RS's profile, that are not immediately apparent to a board member ([e.g.,] the RS has a small profile that consists of only exceptional Marines...).”<sup>21</sup> The problem with this guidance is that the only “conflict” in this case is that an above average Marine will receive a report with an RV of 80. But HQMC does not provide RSs with the means to calculate RVs in advance. Rather, they are told that RVs are “not relevant to the fitness report writing process.”<sup>22</sup> It should be no surprise, then, that RSs often mark their reports without addressing the conflict between the RV and their comments—there is a good chance they don't know it's there.

Perhaps ironically, MMPB says this kind of ambiguity—where the RV does not match the RS comments—is the “most consistent complaint” selection board members have about fitness reports.<sup>23</sup> MMPB argues the problem stems from RSs either deliberately giving Marines the same evaluation or marking a report in a way that ensures it receives a preselected RV or FRA rather than following the guidance in the PES Manual.<sup>24</sup> But that need not be the case. As the example above shows, RVs can diverge from the RS comments even when the RS carefully follows the PES Manual.

One final note about meaningful range: MMPB is aware of this assumption and its importance (they call it “quality spread”). In a brief posted on their website, they note that “our system is dependent upon quality spread. Without quality spread, our current evaluation system is not sustainable.”<sup>25</sup> This is true. Unfortunately, the PES Manual makes no provision for RSs to mark their reports in a way that ensures quality spread. It admits that “RSs who consistently mark all of their Marines the same do a disservice to

<sup>20</sup> For this example, I assume the staff sergeant who received the slightly higher evaluation was transferred.

<sup>21</sup> PES Manual, p. 4-40.

<sup>22</sup> PES Manual, p. 8-6.

<sup>23</sup> PES Brief Part IV, slide 17.

<sup>24</sup> PES Brief Part IV, slide 18.

<sup>25</sup> PES Brief Part IV, slide 10.

their Marines.”<sup>26</sup> But it also emphatically states that “grades are earned by the MRO’s displayed efforts and apparent results; they are not given to attain a perceived fitness report average or relative value.”<sup>27</sup> Current policy, then, requires us to *hope* that the performance of Marines under each RS is diverse enough to ensure meaningful range.<sup>28</sup>

### Assumption 3: Large sample size

The PES assumes every RS has a marking philosophy—a set of marks the RS considers average together with some criteria for giving above or below average marks.<sup>29</sup> There is no requirement to formally document marking philosophies, but the PES assumes they exist and affect the way each RS marks their fitness reports. (The PES Manual discusses marking philosophies in detail beginning on p. 4-21.)

The PES Manual says the RS profile is a “snapshot of the RS’s marking philosophy,” but that language obscures the true relationship between an RS’s profile and their philosophy.<sup>30</sup> RS profiles change over time—the range and average of an RS profile will usually change as the RS writes additional fitness reports. Marking philosophies, by contrast, are fixed.

According to the PES Manual:

Once a RS has decided on a marking philosophy, they must maintain that same marking philosophy for all ranks and all MOSs for the entirety of their career. If a RS were to change their marking philosophy midcareer it would have drastic effects on the Marines for whom they had previously written fitness reports.<sup>31</sup>

The RS profile and the marking philosophy are related—every report in the profile should be consistent with the marking philosophy; however, the profile is both dynamic and available to HQMC, while the marking philosophy is neither.

That said, if the RS marks every report consistent with their marking philosophy, something called the “weak law of large numbers” will ensure the average of the RS profile converges to the average of the marking philosophy as the number of reports increases.<sup>32</sup> When that happens, you can use the average from the RS profile as a stand-in for the marking philosophy average and (literally) subtract out its influence. If you look closely at the equation for RVs, you will see that this is exactly what it tries to do.<sup>33</sup> In other words, the PES does not require RSs to document their marking philosophies

<sup>26</sup> PES Manual, p. 8-5.

<sup>27</sup> PES Manual, p. 4-23. See also p. 4-22, which says, “Attribute grades should be earned, not given; they should reflect the MRO’s exhibited efforts and results; the marks should not be based on a preconceived or artificial fitness report average.” The RS checklist in Appendix F includes similar guidance.

<sup>28</sup> In their educational materials, MMPB says you can calculate the range of the RS profile by subtracting the RS Avg from the RS High and then multiplying the result by two (see, for example, PES Brief Part IV, slide 22). But that’s inaccurate. The calculation returns the range of the RV scale, not the RS profile. The data on the MBS make it possible to calculate the range of the RS profile only in unusual circumstances.

<sup>29</sup> See PES Manual, pp. 4-21 to 4-22.

<sup>30</sup> PES Manual, p. 8-4. NB: “snapshot of the RS’s marking philosophy” is an accurate description of the way RS profiles are *used* in the current system (i.e., the RV equation treats the RS profile as an error-free estimate of the marking philosophy), but it’s not an accurate description of their true relationship.

<sup>31</sup> PES Manual, p. 4-21.

<sup>32</sup> For overviews of the weak law of large numbers (which is not the same thing as the Central Limit Theorem), see Peter M. Aronow and Benjamin T. Miller, *Foundations of Agnostic Statistics* (Cambridge: Cambridge University Press, 2019), 100–1; Kosuke Imai, *Quantitative Social Science: An Introduction* (Princeton: Princeton University Press, 2017), 300–2.

<sup>33</sup> For the RV equation, see Baker and Kwan, “Miscalculating Performance,” 67; Adam Clemens et al., *An Evaluation of the Fitness Report System for Marine Officers*, 8n5.

The problem is that the accuracy of the RV conversion depends on the law of large numbers, and the law works best with large numbers—*much* larger than many RS profiles.

because RVs are supposed to make marking philosophies irrelevant.

The problem is that the accuracy of the RV conversion depends on the law of large numbers, and the law works best with large numbers—*much* larger than many RS profiles. When the number of reports is small, there is more noise than signal, and the profile average will be an unreliable stand-in for the marking philosophy. In these cases, there is simply not enough information in the RS profile to know where a report falls with respect to an RS's marking philosophy.

Demonstrating this fact requires some math. Since the average of the RS profile functions as a numerical estimate of the average of the marking philosophy, we can calculate a confidence interval (or "margin of error") that will give us a sense for how certain we can be that the profile average is a good proxy for the average of the marking philosophy.<sup>34</sup> The calculation involves some assumptions, but in this case, the assumptions strengthen rather than weaken our conclusions because they ensure we're examining a best case scenario for RVs.<sup>35</sup>

How much uncertainty is there? A lot. In a profile with only three reports, **a standard 95 percent confidence interval will always be wider than**

**the entire RV scale.** In other words, the distance between 80 and 100 on the RV scale will always be smaller than the baseline uncertainty about what average performance looks like according to the RS's marking philosophy.

This fact derives from the way the numbers are calculated. By construction, the distance from 80 to 100 on the RV scale is always exactly twice the distance from the profile average to the highest report. Since the mean always falls within the range of the data, the RV scale can never exceed twice the range of the RS profile. By contrast, in a profile with only three reports, the confidence interval will always exceed twice the range for three reasons: (1) the length of the confidence interval is a multiple of the standard error (which is estimated using the standard deviation); (2) the standard deviation of a three-observation sample can never be less than half the range of the data; and (3) even when the standard error takes its lowest possible value, the multiple is large enough to ensure that the length of the confidence interval will be larger than twice the range.<sup>36</sup>

To show this is not mathematical sleight of hand, I used statistical software to randomly generate 100,000 RS profiles, each with three reports. I then

<sup>34</sup> For an accessible discussion of confidence intervals and how to interpret them, see Ethan Bueno de Mesquita and Anthony Fowler, *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis* (Princeton: Princeton University Press, 2022), 102–3.

<sup>35</sup> The calculation assumes the reports in each profile are independent of each other, which is not always the case. When an RS writes multiple reports on the same Marine and those reports are used to calculate RVs, those reports are not independent. In these cases, the correct confidence interval will be even wider.

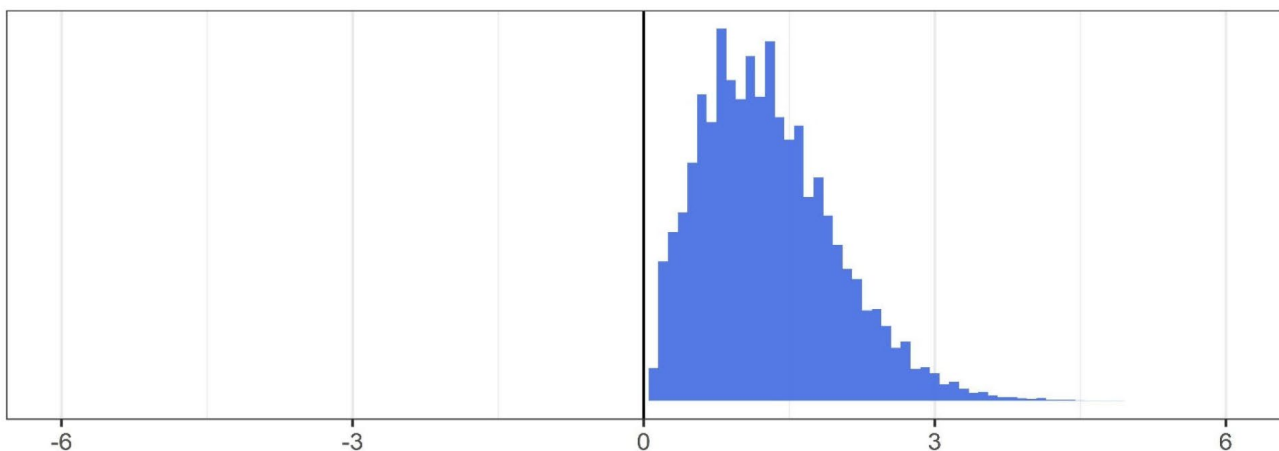
<sup>36</sup> See George W. Thomson, "Bounds for the Ratio of Range to Standard Deviation," *Biometrika* 42, nos. 1–2 (June 1955): 268–69, doi: [10.2307/2333446](https://doi.org/10.2307/2333446). For an overview of how to calculate confidence intervals, see Imai, *Quantitative Social Science*, chap. 7, esp. pp. 339–41.



calculated the range of the RV scale and the range of the 95 percent confidence interval around each profile's average. If the confidence interval is always larger than the RV scale, then when I subtract the length of the confidence interval from the length of the RV scale, the result should always be positive (excluding the corner case when all three reports have the same average, in which case the range of both will be zero). That is exactly what Figure 4 shows. For all 100,000 randomly generated RS profiles with three reports, subtracting the length of the RV scale from the length of the confidence interval always yields a positive number.<sup>37</sup>

In other words, **RVs in small profiles are statistically indistinguishable from random noise**. They are an example of what is sometimes called *false precision*—or, less charitably, *quantifauxcation*: “the common practice of assigning a meaningless number, then concluding that because the result is quantitative, it must mean something (and if the number has six digits of precision, they all matter).”<sup>38</sup> In layman's terms, the margin of error in a poll of three people is much larger than in a poll of 1,000 people—so much larger that the average from the poll of three people is essentially meaningless.

Figure 4



Note: This plot shows the difference between the range of the RV scale and the range of the 95 percent confidence interval in 100,000 randomly generated RS profiles, each with three reports. The entire distribution is to the right of zero because the range of the 95 percent confidence interval is always larger than the range of the RV scale.

<sup>37</sup> The minimum value is 0.1074571. I removed 228 profiles in which all three randomly generated reports had the same average (when that happens, the length of the RV scale and length of the confidence interval are both zero).

<sup>38</sup> Philip B. Stark, “Pay No Attention to the Model Behind the Curtain,” *Pure and Applied Geophysics* 179 (2022), 4122, doi: [10.1007/s00024-022-03137-2](https://doi.org/10.1007/s00024-022-03137-2). As John Tukey warned us nearly 40 years ago: “The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.” See John W. Tukey, “Sunset Salvo,” *American Statistician* 40, no. 1 (Feb. 1986), pp. 74–75, doi: [10.2307/2683137](https://doi.org/10.2307/2683137).

# RVS ARE TOO IMPORTANT TO IGNORE

Does it matter that RVs are ambiguous? They're not the only thing selection boards consider, after all, and they weren't designed to be used in isolation. Perhaps RVs are a net benefit despite their flaws?

This is certainly HQMC's view. In an article for the *Marine Corps Gazette* published in 2016, the former head of the Performance Evaluation Section explained that board members are "not looking for the one mathematical computation to validate a Marine's position" but rather consider "a Marine's totality of service" when making their selections.<sup>39</sup> And in a recent brief on how to read an MBS, HQMC said "RS/RO numbers were not designed to be used in a vacuum. They were developed to give board members a starting point. If you rely solely upon the numbers, you may be misled."<sup>40</sup>

The problem with this view is that it's inconsistent with what we know about the way the Marine Corps uses RVs. We know from members of recent selection boards, for example, that "most briefers immediately look at the RS/RO RVs as soon as the member is shown on [Digital Board Room]. An initial opinion is instantly formed before the briefer has said a word about the Marine they are briefing."<sup>41</sup>



We know from publicly available research that RVs meaningfully affect the probability of selection even when controlling for many other things in a Marine's "totality of service."

We know from publicly available research that RVs meaningfully affect the probability of selection even when controlling for many other things in a Marine's "totality of service," including Professional Military Education, combat experience, personal awards, time in grade, civilian education, commissioning source, physical fitness, occupational field, command experience, and reviewing officer marks.<sup>42</sup> We know from MMPB's educational materials that the "most consistent complaint" board members have about fitness reports is that the RVs don't match

<sup>39</sup> Pritchett, "Rebuttal to 'Miscalculating Performance,'" p. 63.

<sup>40</sup> MMRP, "Master Brief Sheet Overview," slide 25. For a similar view, see Chris Esposito, "It's All Relative: Insights on RO/RS Markings," Career Management Team (CMT) newsletter #11, Sept. 2023.

<sup>41</sup> Scott Clendaniel, "Notes from FY24 LtCol SMCR Lieutenant Colonel Selection Board," (unpublished document, May 2023). See also Vasquez, "FY20 Active Reserve SNCO Board Debrief," (undated PowerPoint presentation), which reveals that 24 percent of board members thought RVs "told the true story" when they diverged from the narrative comments. (Only 29 percent placed more weight on the narrative comments; the rest responded that the two sources of information "were considered equally.")

<sup>42</sup> Joel Hoffman, "Significant Factors in Predicting Promotion to Major, Lieutenant Colonel, and Colonel in the United States Marine Corps" (master's thesis, Naval Postgraduate School, 2008), ch. 5; See also Maj Brian M. Anderson, "The Lieutenant Colonel Command Screening Board: Quantitative Analysis of Career Paths and Selection Results," *Marine Corps Gazette* 107, no. 8 (Aug. 2023): 20–22; Jacob L. Reynolds, "Effect of Being an Aviator on Promotion to O-5 in the USMC" (master's thesis, Naval Postgraduate School, 2011); Raul P. Garza, "United States Marine Corps Career Designation Board: Significant Factors in Predicting Selection" (master's thesis, Naval Postgraduate School, 2014); Shannon Phillips and Adam Clemens, *The Fitness Report System for Marine Officers: Prior Research*, CNA, 2011, CIM D0026273.A1/Final.

the narrative comments, and that “in the absence of an explanation, most members [weigh] the mark and resulting RV more heavily, vice weighing the Section I comments.”<sup>43</sup> And we know that RVs (and other metrics derived from them) are used by monitors to determine who gets choice assignments, as well as by manpower analysts when evaluating the PES, the promotion board process, and a variety of other policies and procedures.<sup>44</sup>

In short, we know that RVs play an important role in the way the PES works and shape important decisions at the service level. Asking boards and analysts not to overrely on them is a good start, but it is clearly not enough to ensure RVs are appropriately discounted in the board room.

---

<sup>43</sup> See, respectively, PES Part IV, slide 17; LtCol Jason W. Heuer, “The Marine Corps Promotion Board Process: An After-Action Report from a Board Member,” *Marine Corps Gazette* 104, no. 12 (Dec. 2020), 50. While there is likely some variation in how boards resolve ambiguity between the narrative comments and RVs, it is enough to show that boards *sometimes* prefer the RVs over other sources of information.

<sup>44</sup> Most of this analysis is not publicly available, but see, for example, William L. Dunst, “Evolution of the Marine Officer Fitness Report: A Multivariate Analysis” (master’s thesis, Naval Postgraduate School, 2018). Regarding monitors, the “7/31/2023 Update” on the webpage for the combat arms monitor for O4s includes a table of the average cumulative RVs of Marines selected for command. The monitor then advises, “If your numbers are well off, and someone else’s is much closer, I will prioritize their assignment to a key billet for their development as a potential future commander.” See <https://www.manpower.usmc.mil/webcenter/portal/OA1CAMaj> (last accessed Oct. 21, 2023).

# RVS MAKE PROHIBITED COMPARISONS

But there is another reason the current system should not be left in place: RVs compare Marines who are not supposed to be compared to one another for promotion purposes.

Each year, the Marine Corps runs separate promotion boards for active and reserve officers. Doing so might seem duplicative, but the separation is required by law. 10 US Code § 611(a) gives the Secretary of the Navy the authority to convene a promotion board for officers on the Active-Duty List (ADL), while 10 US Code § 14101 gives the Secretary the authority to convene one for officers on the Reserve Active-Status List (RASL). The law says that officers can be on only the ADL or the RASL, not both, and that each promotion board can consider officers on only one list or the other, not both.<sup>45</sup>

The legal separation between active and reserve officers has some downsides (most notably for permeability between components), but one upside is that it helps ensure promotion boards compare “apples to apples” when making their recommendations. Promotion is a competitive process, and it makes little sense to compare Marines in full-time active-duty service to Marines in a part-time status.<sup>46</sup>

Unfortunately, this is exactly what RVs do. RSs have a separate profile for each rank but not for each

status. RSs who have written reports on both full- and part-time Marines of the same rank will see those reports combined into a single RS profile and used to calculate RVs and other profile statistics. As a result, **RVs allow promotion boards to compare Marines from different statuses—from both the ADL and RASL—to one another in a way that circumvents the intended separation of active from reserve boards.**<sup>47</sup>

In most cases, these comparisons are both inconsistent and unwitting. They are inconsistent because not every RS has written reports on both full- and part-time Marines. Some part-time Marines will have RVs derived from profiles composed entirely of other part-time Marines, while others will have RVs derived from RS profiles in which they are the only part-time Marine.

These comparisons are unwitting because selection boards have no way to know which Marines are burdened with this kind of “apples to oranges” comparison unless the RS is fastidious enough to mention it in their narrative comments. (And even then, boards have no way to verify the accuracy of the comments or whether they remain accurate after the RS has submitted additional reports.)

By combining active and reserve reports in the same profile and masking the breakdown from selection

<sup>45</sup> See 10 US Code § 14002 for why officers can be on only one list or the other. See §§ 14004, 14102(b), 611(a), and 612 for why promotion boards can consider officers on one list or the other, but not both. For helpful overviews of the procedural requirements for promotion boards, see MCO P1400.31C (the “Promotion Manual”) and RAND’s “DOPMA/ROPMA Policy Reference Tool,” available at <https://www.rand.org/paf/projects/dopma-ropma.html>.

<sup>46</sup> Marines in the Active Reserve program are listed on the RASL but are considered for promotion within a separate competitive category. So, while they are full-time Marines, they do not compete for promotion against their part-time counterparts (except as described in this section). See MCO P1400.31C, p. 1-14.

<sup>47</sup> NB: Separate competitive categories suffer from a similar problem. See [10 USC § 621](#) and [§ 14005](#).

boards, the Marine Corps is unintentionally stacking the deck against a subset of reserve Marines when they go up for promotion. As the 38th Commandant put it in his 2019 planning guidance:

[W]e must accept the realities related to periods of annual training completed by our Reserve Component (RC) Marines. For many, these periods of service require an Active Component Marine to complete a fitness report covering two weeks. Because these reports are weighted the same as every other report within a RS's profile, they are habitually a low relative value to

avoid artificially skewing the RS's profile. While this is understandable, it should not endure. We must provide the RS with an opportunity to evaluate the individual's performance in relation to every other RC Marine the RS has evaluated completing similar training, and not attempt to judge the two week performance against periods usually covering six months—if not longer.<sup>48</sup>

Unfortunately, despite this high-level recognition, active and reserve reports are still combined into a single profile for each rank.

---

<sup>48</sup> David M. Berger, *Commandant's Planning Guidance*, US Marine Corps, 2019, 8.

# A GRAPHICAL APPROACH WOULD BE BETTER

Table 2 provides another view of the ambiguity problem I described earlier. It shows five notional RS profiles, one in each column. As in Table 1, the numbers in the bottom half of the table show how the highlighted report in each column would appear on an MBS. *Unlike* in Table 1, the reports here are listed in rank order rather than the order in which they were processed (hence why there is no value in the "RV at processing" row). Note that the MBS information is identical even as the position of the report within the RS profile changes, with the minor

exception of the number of reports in Profile 5. In this case, it is mathematically impossible to infer the location of the highlighted report within the profile from the information on the MBS alone. There is simply not enough information (except for the report at the top of the profile, which always receives an RV of 100).

While it may not be obvious at first, what you see in the table is an example of a well-understood problem in statistics. As a rule, you lose information

Table 2

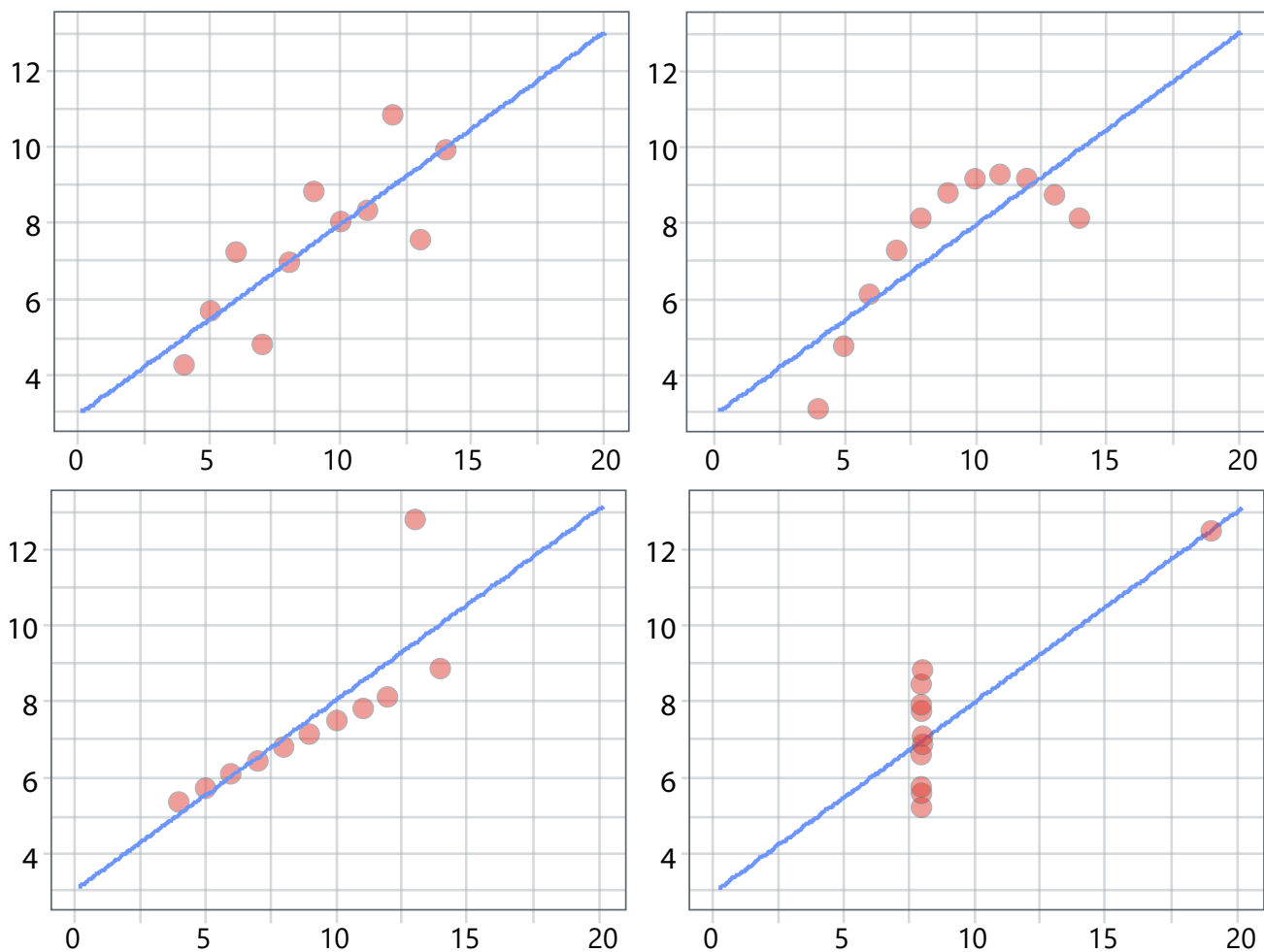
	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
<b>RS Profile</b>					
Report 1	5.00	5.00	5.00	5.00	5.00
Report 2	3.38	3.54	4.00	4.77	<b>3.00</b>
Report 3	3.31	3.46	3.38	<b>3.00</b>	<b>3.00</b>
Report 4	3.23	3.31	<b>3.00</b>	2.77	<b>3.00</b>
Report 5	3.08	<b>3.00</b>	2.92	2.77	
Report 6	<b>3.00</b>	2.69	2.69	2.69	
<b>Master Brief Sheet</b>					
Reports	6	6	6	6	4
RPT Avg	3.00	3.00	3.00	3.00	3.00
RS Avg	3.50	3.50	3.50	3.50	3.50
RS High	5.00	5.00	5.00	5.00	5.00
RPT at High	1	1	1	1	1
RV at Proc	-	-	-	-	-
Cum RV	86.67	86.67	86.67	86.67	86.67

Note: This table shows the disconnect between the RS profile and MBS. The top portion shows five alternative RS profiles. The bottom portion shows how the highlighted reports in each profile are represented on the MBS. RSs can see the top portion; selection boards can see the bottom portion. Neither can see what the other sees.

when you summarize data. You can use the attribute marks on a fitness report to calculate a report average (a summary of the data), but you can't use the report average to figure out what the original attribute marks were.<sup>49</sup> That doesn't mean you shouldn't summarize data, but it's important to keep in mind what the summaries omit.

The way many statistics textbooks make this point is with Anscombe's Quartet, a set of four scatterplots created by the statistician Francis Anscombe in the early 1970s (see Figure 5).<sup>50</sup> The scatterplots each look very different, but they have the same summary statistics—the mean and standard deviation of each plot's x values are the same; the mean and standard deviation of each plot's y values are the same; and

Figure 5



Note: Anscombe's Quartet. The summary statistics for each plot are the same, including the mean of each plot's x values, the mean of each plot's y values, the coefficient of determination, the correlation between x and y, and the linear regression line (in blue).

<sup>49</sup> Except when the report average is at its formal maximum or minimum value, which is likely rare. For more on "summaries of data," see David J. Hand, *Dark Data: Why What You Don't Know Matters* (Princeton: Princeton University Press, 2020), e.g., 297.

<sup>50</sup> F. J. Anscombe, "Graphs in Statistical Analysis," *American Statistician* 27, no. 1 (1973): 17–21, doi: [10.1080/00031305.1973.10478966](https://doi.org/10.1080/00031305.1973.10478966). The data for Anscombe's original plots are included with the free statistical program R (just type "anscombe" into the command prompt).



An RV is a statistical calculation that makes assumptions about the data (in this case, the RS profile). Those assumptions are often wrong, but because selection boards cannot see the underlying data, no one in the board room is the wiser.

the correlation coefficient, number of observations, and regression line in each plot (shown in blue) are also the same.<sup>51</sup> The plots are clearly different, but those differences are undetectable from the summary statistics alone.

Anscombe used these plots to argue that statisticians should look at both calculations and graphs in their analysis: “Most kinds of statistical calculation rest on assumptions about the behavior of the data,” he wrote. “Those assumptions may be false, and then the calculations may be misleading. We ought always to try to check whether the assumptions are reasonably correct; and if they are wrong we ought to be able to perceive in what ways they are wrong. Graphs are very valuable for these purposes.”<sup>52</sup>

This statement precisely captures the problem with the way RVs are presented on MBSs. An RV is a statistical calculation that makes assumptions about the data (in this case, the RS profile). Those assumptions are often wrong, but because selection boards cannot see the underlying data, no one in the board room is the wiser. In fact, no one *at all* is the wiser. The way the current process works, no one is supposed to have access to both the RS profile and the RVs—not the MRO, the RS, the RO, or even the members of a selection board. In essence, the system is set up so that no one is allowed to see both the calculations and the graphs, so misinterpretations go entirely unnoticed.

The simplest solution to this problem is to show selection boards the graphs. As the figures throughout this article demonstrate, a graphical representation of an RS profile is both feasible and does a much better job of revealing the quirks in an RS profile than the summary statistics. In fact, for small profiles, a graphical representation is among the best ways to convey the uncertainty inherent in the performance evaluation process. To that end, Figure 6 shows what a graphical MBS might look like using the five notional RS profiles from Table 2.

These plots are designed to replace the section of the MBS shown earlier in Figure 1. No RVs are listed, of course, but a few other things are missing too, including the mean of the RS profile (“RS Avg”), the number of reports at the top of the profile (“RPT at High”), and the order in which the report was completed and processed by HQMC (“X of Y”). These numbers are missing because they’re no longer necessary.<sup>53</sup> A count of the number of reports at the top of the profile has limited value when the entire profile is visible on the MBS. Likewise with processing order—without the RV at processing to worry about, there is no substantive reason to display the order in which a report was processed by HQMC.

The need to stop using the RS average is less obvious but more important. The RS average (the mean) is sensitive to outliers and makes little sense in small profiles—it is, in fact, one reason RVs are so quirky.

<sup>51</sup> As used here, the word *same* means identical to at least two decimal places.

<sup>52</sup> Anscombe, “Graphs in Statistical Analysis,” 17.

<sup>53</sup> The list of attribute marks is also missing, but not because they are actively harmful. These marks take up a lot of space on the MBS but do not appear to add much value to selection board deliberations. Some board members likely use them to quickly find “F” and “G” marks, but a simple flag for these marks (like the flag for adversity) would be a much more efficient use of space—especially since a board member searching for an “F” or “G” will, if they find one, almost certainly open the corresponding fitness report to read the justification comments.



The median is much less volatile and serves as a more reliable indicator for where a Marine falls in the profile. Half the reports in a profile will always be above the median, and half will always be below. In Figure 6, the median is depicted as a vertical blue line rather than a number, but that's a matter of personal preference. It can be displayed in either format.

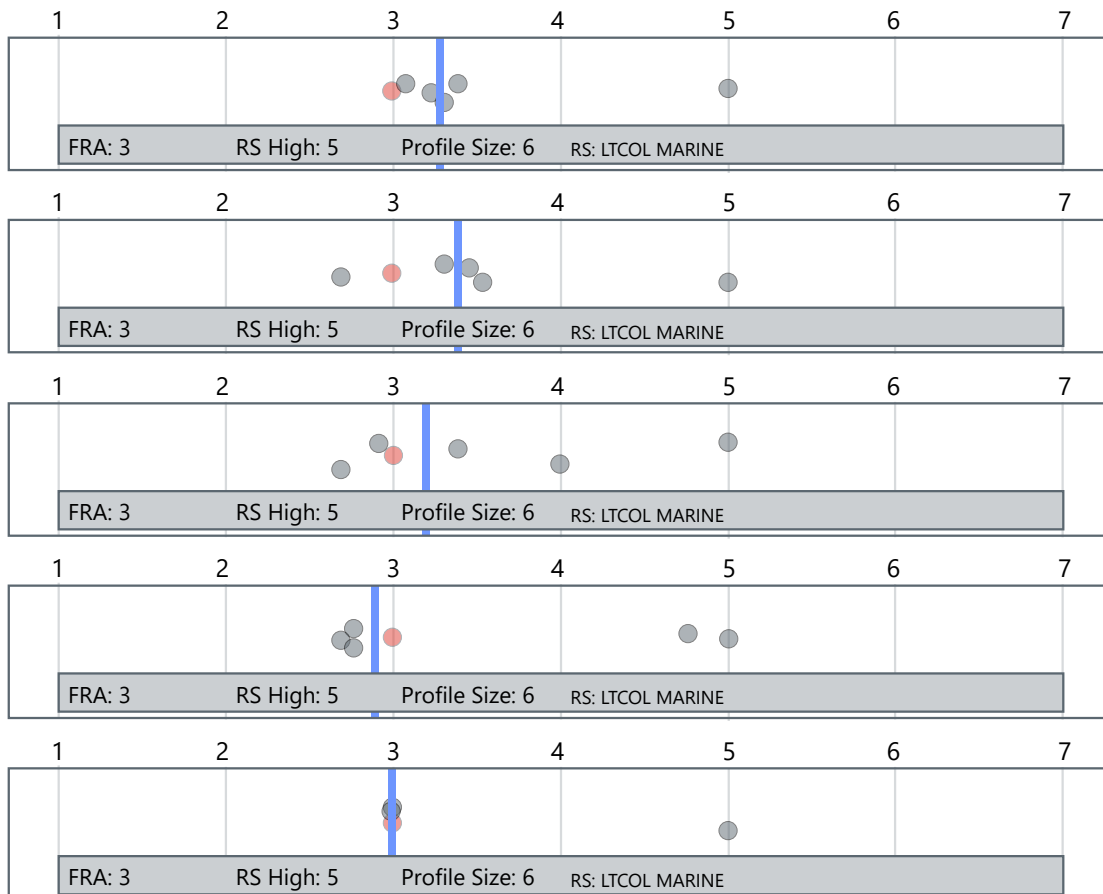
Two final points about this alternative MBS:

1. It's still possible to bin reports into thirds. But now, instead of using an obscure equation to define the threshold between bins, the Marine Corps can use

the same intuitive method it uses at TBS: put an equal number of reports in each third.<sup>54</sup>

2. The five profiles in Figure 6 are the same five profiles presented in Table 2. Here, though, the profiles are easy to distinguish—the summary statistics are still identical (except the median, which I've added), but each profile is visibly different, and the relative performance in each case clearly varies. The marks on the highlighted report are identical in each profile, but the position of the other reports provides valuable context.

Figure 6



Note: This figure depicts what a “graphical MBS” might look like, using the five notional RS profiles from Table 2. As in earlier figures, each dot is one FRA, and a set of dots is one RS profile. The vertical blue line in each plot is the median of the RS profile. The FRA and RS of the highlighted report (the purple dot) are displayed in the information bar in each plot.

<sup>54</sup> You can even display the thirds on the plots. I chose not to because they made the figure look cluttered.

## CONCLUDING REMARKS

Eliminating RVs and putting graphs on MBSs will not solve every problem with PES or the selection board process. It won't do anything to help board members struggling to weigh reports from critical billets, account for multiple reports from the same RS, or reliably compare 90-day to 365-day periods of observation. Nor will it make it easier for RSs to write narrative comments that clarify rather than obscure their true evaluation. But it will do one very important thing: it will prevent selection boards from being actively misled by a mathematical quirk.

RVs are unique among the known problems with the PES in that they involve formal assumptions that few

understand and no one is allowed to verify. Most other problems with the PES are straightforward and well understood by members of the board—they know, for example, that a two-week reserve training (RT) report isn't the same thing as a 12-month grade change (GC) report. The MBS also includes all the information the board needs to have a productive deliberation about how to compare the two. The same cannot be said for RVs. They are a wild card that corrupts the process, and the Marine Corps should stop using them.

# APPENDIX A

An astute member of a selection board can detect when there is a problem with meaningful range by paying attention to the gap between the RS average and the RS high on an MBS. If the board member is aware that a difference of about 0.07 on the report average scale roughly equates to one attribute mark, they can convert the range of the RV scale back into the attribute mark scale by subtracting the RS average from the RS high, multiplying by 2, and then dividing by 0.07.

In equation form:

$$\frac{2(RS\ high - RS\ avg)}{0.07}$$

This number will be an estimate of the number of attribute marks that separate the top and bottom of the RV scale (note: the range of the RV scale is not the same as the range of the RS profile, as discussed in the main text).<sup>55</sup> The smaller the number, the smaller the range and the more cautious the board should be. This method is imperfect—if the RS has given out a lot of “non-observed” attribute marks (box “H”), the correct conversion factor will be larger than 0.07, but if the calculation returns a very low number using 0.07, the RS profile very likely does not have enough range for RVs to correspond to meaningful differences in performance.

---

<sup>55</sup> In their educational materials, MMPB says you can calculate the range of the RS profile by subtracting the RS Avg from the RS High and then multiplying the result by two (see, for example, PES Brief Part IV, slide 22). But that’s inaccurate. The calculation returns the range of the RV scale, not the RS profile. The data on the MBS make it possible to calculate the range of the RS profile only in unusual circumstances.

## APPENDIX B

I relied on R and several R packages while researching this article. I note them here.<sup>56</sup>

Allaire, J.J., Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2023. *Rmarkdown: Dynamic Documents for R*. <https://github.com/rstudio/rmarkdown>.

Hester, Jim, and Jennifer Bryan. 2022. *Glue: Interpreted String Literals*. <https://CRAN.R-project.org/package=glue>.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, and JooYoung Seo. 2023. *Gt: Easily Create Presentation-Ready Display Tables*. <https://CRAN.R-project.org/package=gt>.

Mersmann, Olaf, Heike Trautmann, Detlef Steuer, and Björn Bornkamp. 2023. *Truncnorm: Truncated Normal Distribution*. <https://CRAN.R-project.org/package=truncnorm>.

Pedersen, Thomas Lin. 2023. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Boca Raton, FL: Chapman & Hall/CRC, 2014.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC. <https://yihui.org/knitr/>.

———. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.

Xie, Yihui, J. J. Allaire, and Garrett Golemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, FL: Chapman & Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, FL: Chapman & Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>.

<sup>56</sup> On the importance of citing software, see Arel-Bundock and McCrain, "Software Citations in Political Science," *PS: Political Science & Politics* 56, no. 3 (2023): 398–401, doi: [10.1017/S1049096523000239](https://doi.org/10.1017/S1049096523000239).

**This report was written by CNA's Resources and Force Readiness Division (RFR).**

RFR provides analytic support grounded in data to inform resource, process, and policy decisions that affect military and force readiness. RFR's quantitative and qualitative analyses provide insights on a full range of resource allocation and investment decisions, including those pertaining to manning, maintenance, supply, and training. Drawing on years of accumulated individual and unit data, as well as primary data collections, the RFR toolbox includes predictive data analytics, statistical analysis, and simulation to answer optimization and what-if questions, allowing military leaders to make better informed decisions.

Any copyright in this work is subject to the Government's Unlimited Rights license as defined in DFARS 252.227-7013 and/or DFARS 252.227-7014. The reproduction of this work for commercial purposes is strictly prohibited. Nongovernmental users may copy and distribute this document noncommercially, in any medium, provided that the copyright notice is reproduced in all copies. Nongovernmental users may not use technical measures to obstruct or control the reading or further copying of the copies they make or distribute. Nongovernmental users may not accept compensation of any manner in exchange for copies.

All other rights reserved. The provision of this data and/or source code is without warranties or guarantees to the Recipient Party by the Supplying Party with respect to the intended use of the supplied information. Nor shall the Supplying Party be liable to the Recipient Party for any errors or omissions in the supplied information.

This report may contain hyperlinks to websites and servers maintained by third parties. CNA does not control, evaluate, endorse, or guarantee content found in those sites. We do not assume any responsibility or liability for the actions, products, services, and content of those sites or the parties that operate them.

Various laws, regulations, and legal issues are discussed in this report. The findings and analysis in this report are for informational purposes only and do not provide legal advice. Do not act on, or refrain from acting on, what you may perceive to be legal advice or a legal conclusion based on anything you read in this report. Seek the assistance of a licensed attorney for any matters you regard as requiring legal analysis or a legal opinion.

## **ABOUT CNA**

CNA is a nonprofit research and analysis organization dedicated to the safety and security of the nation. It operates the Center for Naval Analyses—the federally funded research and development center (FFRDC) of the Department of the Navy—as well as the Institute for Public Research. CNA develops actionable solutions to complex problems of national importance. With nearly 700 scientists, analysts, and professional staff, CNA takes a real-world approach to gathering data. Its unique Field Program places analysts on aircraft carriers and military bases, in squad rooms and crisis centers, working side by side with operators and decision-makers around the world. CNA supports naval operations, fleet readiness, and strategic competition. Its non-defense research portfolio includes criminal justice, homeland security, and data management.



Dedicated to the Safety and Security of the Nation

[www.cna.org](http://www.cna.org)