VALTERA®

The Power of Clarity™

WHITE PAPER

Weighting Survey Results

Methods for Applying Weights to Employee Opinion Survey Results

Scott A. Young, Ph.D.

When surveying an organization (or group), the ultimate goal is to understand the attitudes of the entire organization (or group). This is simple in a perfect world where everyone can be invited to participate and a 100% response rate is achieved. However, because of practical considerations (e.g., cost and employees failing to respond) participation rates rarely approach 100%. As a result, an incomplete or inaccurate picture of the larger group or organization may emerge when only a sample of the organization completes the survey. Item weighting is one method to address this potential problem.

This white paper describes when item weighting is appropriate, and contains detailed instructions on how to calculate and apply the correct weights. Numerous situations and weighting schemes are presented.

About Valtera

Valtera is an HR consulting firm that helps global companies align their people and service delivery to their core business strategy.

Contact

T: +1 847 640 8820
F: +1 847 640 8830
sales@valtera.com
www.valtera.com

CHICAGO CORPORATE OFFICE

Valtera Corporation
Continental Towers
1701 Golf Road, Suite 2-1100
Rolling Meadows, IL 60008

Why Use Weighting?

Item weighting may be appropriate for data analysis and reporting whenever the proportions of survey respondents for your groups do not match their proportions in the overall populations of interest. Imagine, for example, the following situation:

A company surveys three of its departments (A, B, and C). Together, these three departments employ 5,000 people. Each department represents between 20% to 40% of the population (see *Figure 1*). Only 500 employees responded to the survey, with participation rates varying by department. For example, though Department A represents 40% of the population, it only represents 20% of the respondents. Contrast this with Department C, which also represents 40% of the population, but 60% of the respondents.

Assume that we simply combine all respondents (with no weighting) in the calculation of the overall results. In this case, Department C would have much more influence than Department A on the results, despite their equal representation in the population. Now assume that employees in Department C are much more satisfied on average than employees from Department A or Department B. We would then end up with a higher proportion of satisfied employees in our sample than actually exists among the population of these three departments. Therefore, this would probably be a good time to use weighting to help ensure that our overall picture is as accurate as possible.

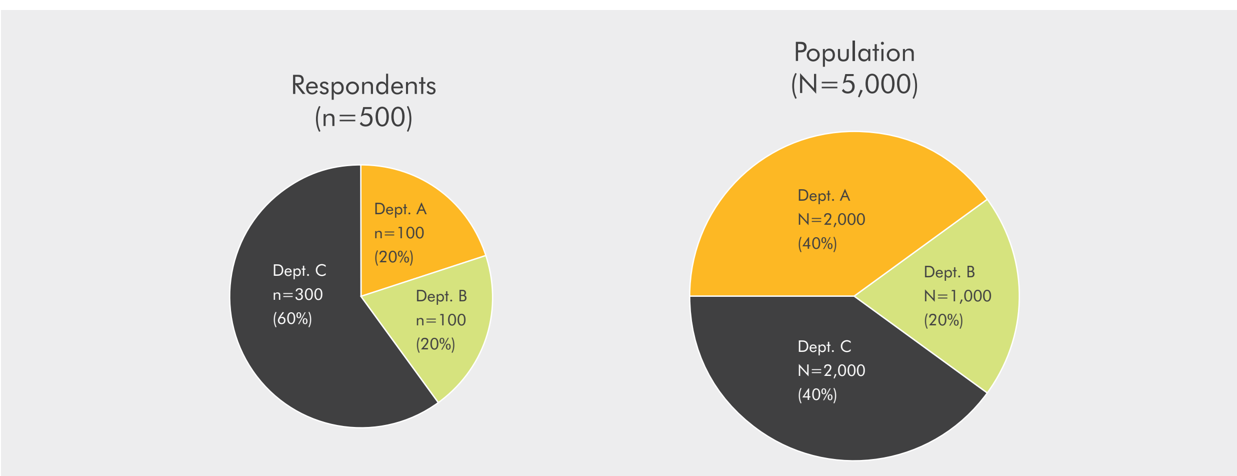
Reasons for Weighting

There are two primary reasons why the proportion of respondents might differ from proportions in the overall population.

Sampling Techniques May Require Different Proportions from Certain Groups

When data are to be reported for different groups (e.g., job levels, departments, etc.), it is important to ensure that these group-level results can be reported with an acceptable level of precision (see sidebar on next page). Higher participation rates are needed in smaller groups than larger groups to obtain a given level of precision for that group. For example, an 80% participation rate is needed for a group of 100 employees to achieve the same level of precision obtained with only a 28% participation rate in a group of 1000 employees. Therefore, higher percentages of small groups are often sampled, relative

Figure 1



to large groups. However, if the data from these two aforementioned groups are also aggregated to determine the overall organizational results, the smaller group would be over-represented relative to the larger group (the smaller group composes 26% of the sample but only 9% of the total organization).

Different groups may have different return rates

Even in the case where equal percentages of employees from each group are invited to participate, actual group response rates may differ for many reasons. For example, one department may have an unusually high response rate because its management communicates strong support for the survey process. In addition, survey administration methods may differ between departments, with some methods yielding greater participation than others (e.g., group administration during work hours typically generates more participation than individuals completing the survey on their own time). Weighting helps to "smooth over" or compensate for some of these differences by allowing the individual responses from under-represented groups (Department A, in *Figure 1*) to contribute more to the overall results than the individual responses from over-represented groups (Department C, in *Figure 1*). In other words, the response of an individual from an under-represented group is given a greater weight than a response from an individual from an over-represented group.

Precision refers to the degree of accuracy with which the sample data reflect the entire population. If respondents are sampled randomly, precision can be calculated from the population size and sample size. For example, assuming a population of one million people, and that results are reported in terms of percentages (e.g., percentage favorable), a sample size of 384 would yield a precision of approximately 95%.

Smaller groups require higher participation rates for a given level of precision and thus may be over-represented in the overall results.

CALCULATING & APPLYING WEIGHTS

Calculating the Weights

The calculation of weights is simple, and remains the same regardless of the number of variables on which the weights are based. The basic formula for computing a weight is:

FORMULA

$$\frac{\text{Cell Population / Total Population}}{\text{Cell Returns / Total Returns}}$$

In the example shown in *Figure 1*, the cell is simply the department of interest. Below are the weight calculations for the three departments in the example.

DEPT. A

$$\frac{2{,}000 \;/\; 5{,}000}{100 \;/\; 500} = 2.0$$

DEPT. B

$$\frac{1{,}000 \;/\; 5{,}000}{100 \;/\; 500} = 1.0$$

DEPT. C

$$\frac{2{,}000 \;/\; 5{,}000}{300 \;/\; 500} = 0.67$$

Notice that the responses from Department A are given a large weight (weight=2.0). This compensates for the lower proportion of returns from Department A compared to the proportion of Department A employees in the population. In a similar manner, responses from Department C are given less weight (weight=0.67) while Department B responses are unit weighted (weight=1.0). This process ensures that the contribution of each group corresponds to its representation in the population.

Hint: Checking Your Work

To ensure that the weights were calculated correctly, multiply the returns for each group by their weights and sum the products. The result should be equal to the total number of returns as shown below:

DEPT. A DEPT. B DEPT. C

(2.0 x 100) + (1.0 x 100) + (0.67 x 300) = 500

Applying the Weights

There are two ways to calculate the overall weighted average response for a survey item. The first method involves applying a weight to the response of each individual. First, multiply each individual's response by the weight corresponding to his or her group. Then sum these products across all individuals, and divide by the total number of respondents. The second option is to multiply the item mean (i.e., average) for each group by its proportion in the population, and sum the resulting products.

To illustrate the second method, assume that the average response for a particular item for Departments A, B, and C, are 3.50, 2.75, and 3.40, respectively. The overall item score is then calculated as follows:

DEPT. A DEPT. B DEPT. C

(3.5 x 0.4) + (2.75 x 0.2) + (3.40 x 0.4) = 3.31

VALTERA®

Contending with Missing Demographic Information

A simple weighting example assumes that you know all of the demographic information needed to calculate the weights. Often though, this information is unknown for some respondents. Imagine that department information is missing for 60 of the 500 respondents in our earlier example (see *Figure 2*).

The best solution to this problem is to remove the missing cases from the sampling problem. The total number of returns is now considered to be 440 instead of 500. The total population, of course, does not change – the missing cases have no bearing on the number of people in the organization. The new weights become:

DEPT. A

$$\frac{2{,}000 / 5{,}000}{80 / 440} = 2.2$$

DEPT. B

$$\frac{1{,}000 / 5{,}000}{80 / 440} = 1.1$$

DEPT. C

$$\frac{2{,}000 / 5{,}000}{280 / 440} = 0.63$$

The missing cases are assigned a weight of 1.0, allowing the total number of weighted cases to still sum to 500 as verified below:

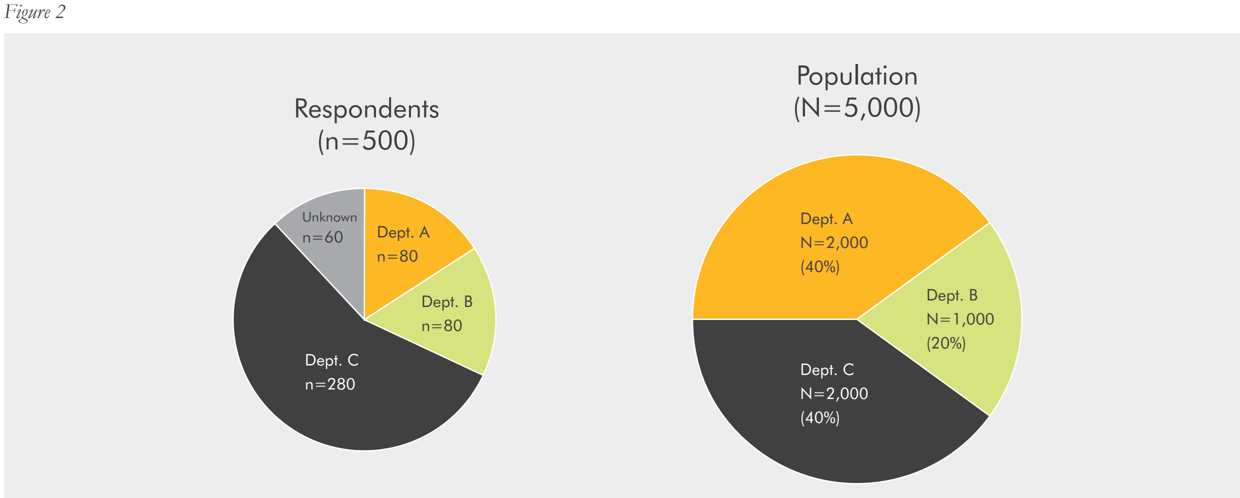
DEPT. A DEPT. B DEPT. C UNKNOWN

(2.2 x 80) + (1.1 x 80) + (0.63 x 280) + (1.0 x 60) = 500

Groups with No Returns

Occasionally, you may have no respondents from a particular group. Most often, it will be because that group was not sampled. In such cases, the group with no returns is simply removed altogether from the sampling problem, which requires subtracting this group's population from the total population. Regardless of whether any employees from this group were sent a survey, that group is left out of our weight calculations—even if they were sent surveys, as you cannot assign weights to non-existent returns!

Figure 2



Multiple Classification Variables

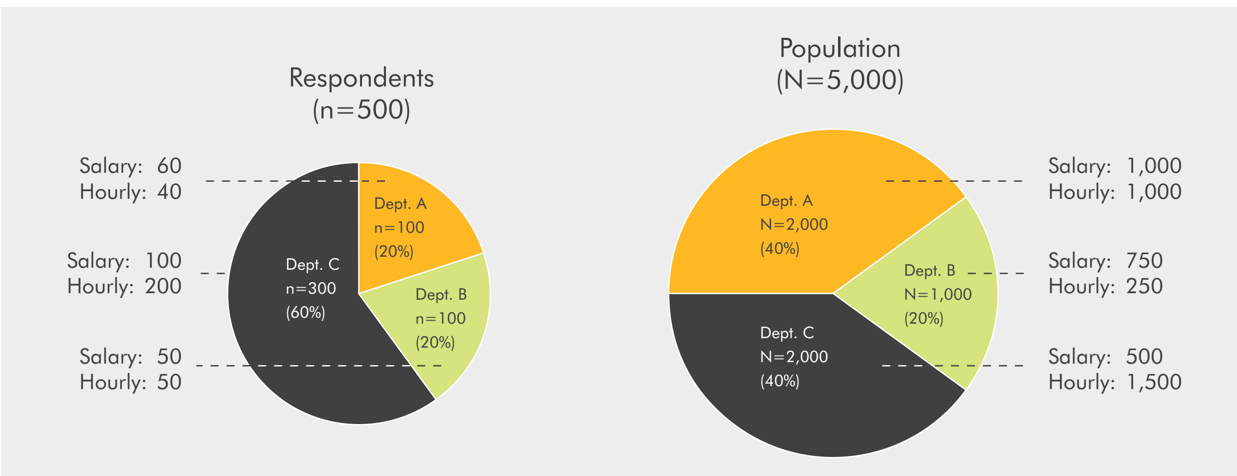
It is often desirable to weight based on multiple variables. For example, in addition to reporting results separately for each department, many organizations compare the results of different pay groups. Therefore, respondents provide both department and pay group information, making it possible to weight the data based on both variables. In this situation, the same basic strategy is applied, with the exception that groups are divided based on both variables together.

Let's return again to our example of Departments A, B, and C. However, now assume we have both salary and hourly employees in those departments. This gives us 3 x 2 = 6 weights to calculate (see *Figure 3*).

As was the case when weighting based on only one variable, we can check our weights by summing the products of the weights and the corresponding number of returns to arrive at the total number of returns:

DEPT. A	(2.5 x 40)	+	(1.67 x 60)	=	200
	+				
DEPT. B	(0.5 x 50)	+	(1.5 x 50)	=	100
	+				
DEPT. C	(0.75 x 200)	+	(0.5 x 100)	=	200
TOTAL					500

Figure 3



	DEPT. A	DEPT. B	DEPT. C
Hourly	1,000 / 5,000 = 2.5 40 / 500	250 / 5,000 = 0.5 50 / 500	1,500 / 5,000 = 0.75 200 / 500
Salary	1,000 / 5,000 = 1.67 60 / 500	750 / 5,000 = 1.5 50 / 500	500 / 5,000 = 0.5 100 / 500

One of the most commonly asked questions is whether it is necessary to compute multiple sets of weights to display the results separately for each pay group, and then to display them separately for each department. The answer is "no." The weights calculated based on both pay group and department can be used to present results for pay group only or department only. Although the results are being presented for only one demographic variable, calculating the weights based on both demographic variables is more appropriate, as it ensures that no department is over-represented in the results for a particular pay group.

Partial Information on Cases

A different problem arises when you are weighting cases based on more than one variable and you have partial information on some cases. Consider an example where 60 cases have missing information for department and/or pay group. Two strategies for dealing with this situation are to apply unit (i.e., 1.0) weights to all cases with missing data, or to compute weights for cases with partial information.

Option 1: Unit Weighting
The first option involves subtracting all of the cases without complete information from the total returns, and calculating weights for the remaining cells based on this new total. In our example, we would base our weight calculations on a total return of 440 cases (500 – 60).

Weight calculations follow the same formula as before using 440 as the number of total returns. And, when you add the unit weighted missing data cases, the products again sum to the total of 500 returns.

DEPT. A	(2.933 x 30)	+	(1.76 x 50)	=	176
+					
DEPT. B	(0.55 x 40)	+	(1.65 x 40)	=	88
+					
DEPT. C	(0.695 x 190)	+	(0.489 x 90)	=	175
+					
UNKNOWN			(1.0 x 60)	=	60
TOTAL					500*

	DEPT. A	DEPT. B	DEPT. C
Hourly	1,000 / 5,000 = 2.933 30 / 440	250 / 5,000 = 0.55 40 / 440	1,500 / 5,000 = 0.695 190 / 440
Salary	1,000 / 5,000 = 1.76 50 / 440	750 / 5,000 = 1.65 40 / 440	500 / 5,000 = 0.489 90 / 440

** Rounding convention used throughout this example leads to a slight variation from the 500 total.*

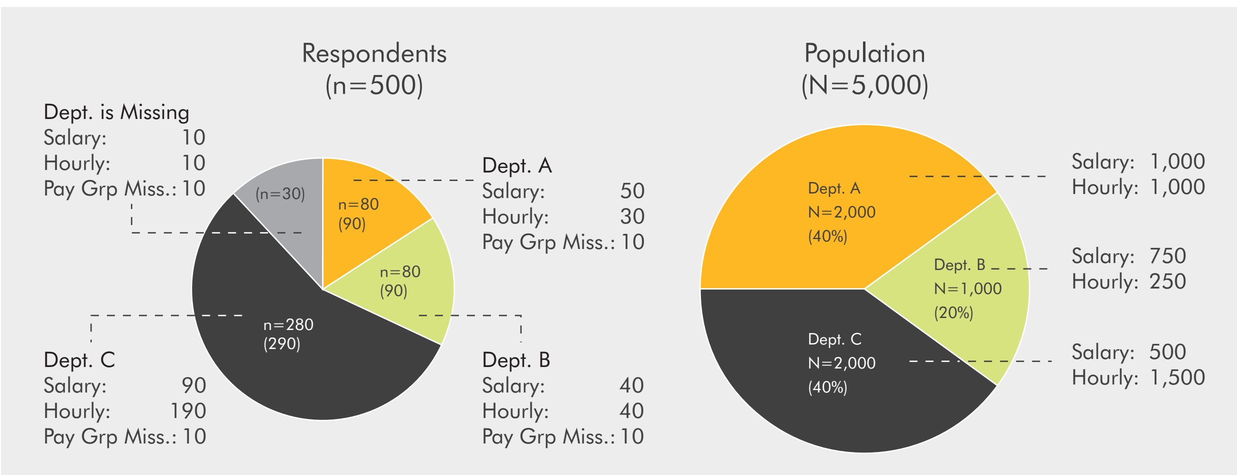
Option 2: Weighting Based on Partial Information
An alternative to unit weighting all cases with any missing data is to weight cases with partially missing information (i.e., missing department or pay group information, but not both) as separate sampling problems. For example, we could weight the cases missing department information based on the numbers of returned surveys in each category of pay group information. Before illustrating, there are a couple of things to remember about this method:

- Each type of missing information (e.g., department missing or pay group missing) is handled separately and independently.
- Cases that have no category information are still unit weighted.
- Using these alternate weights for cases with missing information has no impact on the weights for cases with full information.
- Neither approach is always more appropriate and it will depend on the situation. Some general guidelines are provided below, following another example.

To illustrate and extend our example further, let's assume that of our original 60 cases where department and/or pay group were missing, only 10 have missing data for both variables. The remaining 50 cases with partially missing information are distributed among the remaining categories as shown in *Figure 4*.

In the example, 60 total cases are missing information; 50 are missing one variable, 10 are missing both variables. Notice that within the pie, the sum of the numbers in parentheses is 500—the total number of returns. The sum of numbers with no parentheses is 440—the number of returns with complete information. Again, remember that the weights for these 440 cases with complete information are not affected by the method selected for weighting cases with missing information, and thus, remain unchanged from before.

Figure 4



	DEPT. A	DEPT. B	DEPT. C
Hourly	1,000 / 5,000 = 2.933 30 / 440	250 / 5,000 = 0.55 40 / 440	1,500 / 5,000 = 0.695 190 / 440
Salary	1,000 / 5,000 = 1.76 50 / 440	750 / 5,000 = 1.65 40 / 440	500 / 5,000 = 0.489 90 / 440

Only the 60 cases with missing information are affected. Again, we treat each type of missing information independently, as a separate weighting problem. The formula we have been using is slightly modified. You can see for the cases where pay group information is missing, for example, that 2,750 represents the total number of hourly employees in the population.

FORMULA

DEPARTMENT MISSING

$$\frac{\text{Specific Pay Group Population} \;/\; \text{Total Population}}{\text{Returns from a specific Pay Grp. that are missing Dept. Variable} \;/\; \text{Returns from all known Pay Grps. that are missing Dept. Variable}}$$

HOURLY

$$\frac{2{,}750 \;/\; 5{,}000}{10 \;/\; 20} = 1.1$$

SALARY

$$\frac{2{,}250 \;/\; 5{,}000}{10 \;/\; 20} = 0.9$$

FORMULA

DEPARTMENT MISSING

$$\frac{\text{Specific Pay Department Population} \;/\; \text{Total Population}}{\text{Returns from a specific Dept. that are missing Pay Grp. Variable} \;/\; \text{Returns from all known Depts. that are missing Pay Grp. Variable}}$$

DEPT. A

$$\frac{2{,}000 \;/\; 5{,}000}{10 \;/\; 30} = 1.2$$

DEPT. B

$$\frac{1{,}000 \;/\; 5{,}000}{10 \;/\; 30} = 0.6$$

DEPT. C

$$\frac{2{,}000 \;/\; 5{,}000}{10 \;/\; 30} = 1.2$$

Again, the weights multiplied by the number of returns sums to 500 (the cases where no information is missing sums to 500 – 60 = 440). Notice that we include the unit weights for the 10 cases where both department and pay group information are missing.

WEIGHTING BASED ON PARTIAL INFORMATION

DEPT. A	(2.93 x 30)	+	(1.76 x 50) =	176
+				
DEPT. B	(0.55 x 40)	+	(1.65 x 40) =	88
+				
DEPT. C	(0.69 x 190)	+	(0.49 x 90) =	175
TOTAL				440

CASES WITH MISSING INFORMATION

PAY GROUP KNOWN, NO DEPT.

HOURLY	(1.1 x 10)	=	10
+			
SALARY	(0.9 x 10)	=	10

DEPT. KNOWN, NO PAY GROUP

DEPT. A	(1.2 x 10)	=	12
+			
DEPT. B	(0.6 x 10)	=	6
+			
DEPT. C	(1.2 x 10)	=	12
DEPT. & PAY GROUP MISSING	(1.0 x 10)	=	10
TOTAL			60

WEIGHTING BASED ON PARTIAL INFORMATION	440
+	
DEPT. AND/OR PAY GROUP MISSING	60
TOTAL	500

Which of the two solutions is most appropriate? Is it more appropriate to calculate weights for cases with partial information, or to assign a weight of 1.0 to all incomplete cases? This in part depends on the number of missing responses, and the degree to which the groups involved have return rates disproportionate to their relative group size in the population. The ideal approach is to calculate the weights both ways and determine whether either solution yields group weights that are grossly out of line with the weights of similar groups.

For example, when the missing cases are treated as a separate sampling problem, a situation could arise where the hourly respondents with missing department information count for nearly twice as much as any of the other hourly respondents. In this case, it is preferable to simply assign all of the missing cases weights of 1.0. However, if the return rate for the overall salaried group were 2 to 3 times that of the overall hourly group, it would make more sense to weight the missing cases. If they are not weighted, the hourly respondents with missing department information would each be unit weighted (1.0), while the other hourly respondents would likely have weights of 2 or more.

It is best to apply weights to a group only when the sample from the group is large enough to ensure an accurate representation of the group's opinions.

When is it Inappropriate to Weight Data?

At first glance it appears that data should be weighted whenever the sample proportions do not perfectly represent the organization as a whole. However, there are some circumstances where weighting is not advisable. For example, when return rates are highly disproportionate and the number of returns is very low for certain groups, weighting the data can result in a severe reduction in the precision of the final results. Consider the example below (*Figure 5*).

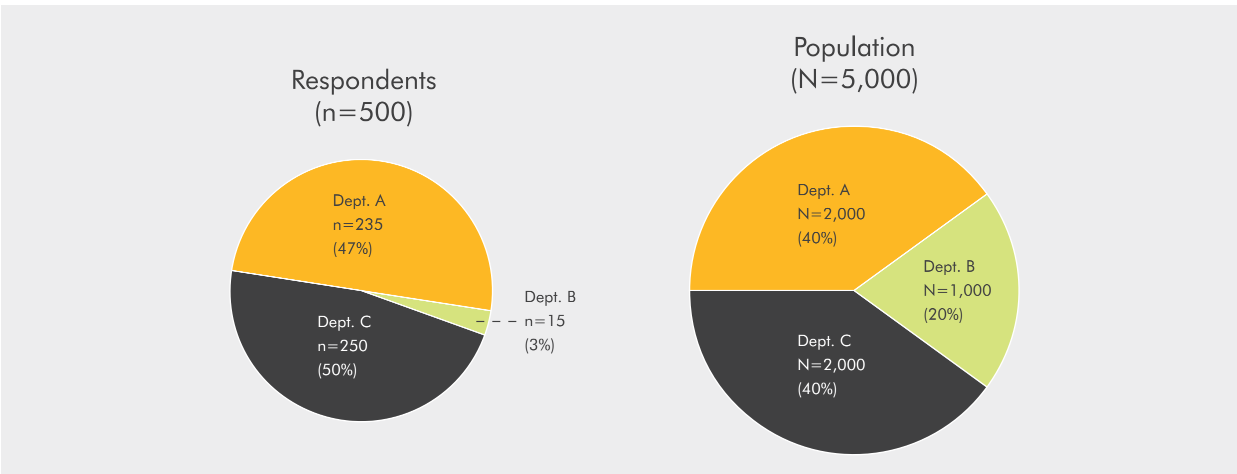
Only 15 people (3%) from Department B responded to the survey. If weights are applied to each department, respondents from Department B are assigned a weight of 6.67. This means that only 3% of the respondents are contributing 20% of the overall results.

DEPT. B

$$\frac{(6.67 \times 15)}{500} = 0.20$$

It is unlikely that the opinions of 15 people will accurately reflect those of the department as a whole. In this case, weighting serves to compound the problem, by giving even more weight to already questionable data. As a general rule, it is best to apply weights to a group only when the sample from the group is large enough to ensure an accurate representation of the group's opinions.

Figure 5



The Power of Clarity.™

Valtera Corporation
Continental Towers
1701 Golf Road, 2-1100
Rolling Meadows, IL 60008

T: +1 847 640 8820
F: +1 847 640 8830

Southern Region
Valtera Corporation
601 Jefferson Street, Ste. 3900
Houston, TX 77002

T: +1 713 650 6535
F: +1 713 650 6595

Southeast Region
Valtera Corporation
301 N. Main Street, Ste. 1103
Greenville, SC 29601

T: +1 864 527 5956
F: +1 864 240 8434

West Coast Region
Valtera Corporation
Xerox Centre
1851 E. First Street, Ste. 900
Santa Ana, CA 92705

T: +1 714 619 9363
F: +1 206 600 6228

East Coast Region
Valtera Corporation
1250 Connecticut Ave., NW, Ste. 200
Washington, DC 20036

T: +1 202 261 6593
F: +1 202 261 6556

The Netherlands
Valtera Netherlands B.V.
Prof. W.H. Keesomlaan 10
1183 DJ Amstelveen

T: +31 (20) 30 50 417
F: +31 (20) 30 50 410

Czech Republic
Valtera CZ s.r.o.
Kremencova 18
Prague 1, 110 00

T: 011-420-234-116-152
F: 011-420-234-116-122