1) Why can we do principal components even if we don't have the raw data?

The principal components are equal to the eigenvectors from either the covariance matrix or the correlation matrix from the data set. Since all you need to find the eigenvectors is the covariance or correlation, that's all you need to find the principal components.

The reason this works is that basically the covariance/correlation matrix describes the data's "original axis" and that's all we need to know in order to figure out how to rotate it into a new axis system.

2) Give the formula for determining the fifth principal component in terms of the original standardized variables.

We can read this directly from the "loadings" for the fifth component.

```
          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
Classics -0.462                -0.128  0.266  0.835
French   -0.441 -0.119         0.228  0.734 -0.448
English  -0.416        -0.342 -0.765 -0.191 -0.298
Math     -0.397  0.255 -0.558  0.567 -0.380
Pitch    -0.367 -0.712  0.388  0.157 -0.425
Music    -0.356  0.643  0.648        -0.171
```

So principal component 5 = 0.266 standardized Classics score + 0.734 standardized French Score
        -0.191 stand. English Score – 0.380 stand. Math score – 0.425 stand. Pitch - 0.171 stand. Music

3) Give the mean, variance, and standard deviation of the students' fifth principal component.

According to the summary, the standard deviation is 0.52000750, squaring that gives a variance of 0.27. Traditionally the principal components are set to have mean 0, and that's what R and SAS do.

```
> summary(spear.pca)
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     2.0255297 0.7868352 0.7153546 0.59753636 0.52000750
```

4) The correlation between the students' standardized classical and French scores scores is 0.83. Give the correlation between the students' first two principal components.

Zero!

5) Some of the coefficients are not shown in the output because their values were smaller than 0.10. What could we use (e.g. ask R for) to find these missing values?

Find the eigen vectors. The function in R is eigen().

6) Briefly interpret the first three principal components.

From the loading matrix above, the first seems to be "overall lack of ability". In order to have a high first component, you need to be low on all of the other standardized abilities on average.

The second is basically -0.712 Pitch + 0.643 Music. To score high on this you would need to be "good at music as opposed to pitch". A low score would mean you were good at telling pitch but not good at music.

Using logic similar to the second, the third component is "good at music as opposed to math".