1. independent variable consists of two naturally occurring groups (e.g., males/females) (1 pt), dependent variable consists of something worth measuring (e.g., some personality trait) (1 pt)
   No (1 pt) [the groups were not based on random assignment, so extraneous variables might account for any differences that are found]

2. False (1 pt) \[\sigma_M \text{ is the standard deviation of the distribution of sample means; the standard deviation of the distribution of differences between means is } \sigma_{M_1-M_2}\]

3. (a) 1.03 (1 pt) [The standardized effect size is equal to the difference between condition means divided by the pooled estimate of standard deviation; \(SS_1 = (14)4.3^2 = 258.86; SS_2 = (14)3.8^2 = 202.16; s_p^2 = (258.86 + 202.16)/(14 + 14) = 16.465; s_p = \sqrt{16.465} = 4.058; d = (12.4 - 8.2)/4.058 = 1.03]

   (b) \(t = 2.83\) (2 pts) \[s_{M_1-M_2} = \sqrt{[(16.465/15)+ (16.465/15)]} = 1.482; t = (12.4 - 8.2)/1.482 = 2.834\]

   (c) sample variances are similar, so pooling is allowed (1 pt)
   (d) critical \(t = 1.701\) (1 pt) [based on \(df = 28\)]
   (e) Imitation training improves self-care skills in mentally disabled children (1 pt)

4. 181 (or 180) (2 pts) [\(\delta \text{ value associated with power of .6 and } \alpha = .05 \text{ for a directional test is } 1.90; \text{ for an independent-samples } t \text{ test, } n = 2(1.90/.2)^2 = 180.5 => 181 \text{ per group}\]

5. 11 (1 pt) [homogeneity of variance does not hold here because one group's variance is more than four times the size of the other group's variance; in that case, the \(df\) we use is the \(df\) for the smaller of the two groups (12 – 1 = 11)]

6. use sample sizes of at least 30 (1 pt) [by using large enough sample sizes, the distributions of sample means will be approximately normal, even if the distributions of raw scores are not normally distributed; with approximately normal distributions of sample means, the distribution of differences between sample means will be approximately normal]

7. A (1 pt) [For A, the regions of rejection are the upper and lower .025 of the \(t\) distribution; for B, the region of rejection is the most extreme .01 of the \(t\) distribution in the end dictated by the direction of the alternative hypothesis; thus, B requires a more extreme \(t\) ratio for significance and is therefore less likely to get a significant result; see the diagram below.]
8. **.008** (1 pt) [The probability of obtaining a $t$ ratio as extreme as or more extreme than the observed $t$ at either end of the distribution (i.e., $\pm 2.574$) is the 2-tailed $p$ value given in the R output, namely, .016. For a $t$ value of $-2.574$ or less, we consider only the lower end of the distribution, which contains half of .016, or .008.]

9. **power = .36** (4 pts) [critical $z = \pm 1.96$; $\sigma_M = 2.5$; critical $M = 60 \pm 2.5(1.96) = 55.10, 64.90$; so the regions of rejection are sample means that are less than or equal to 55.10 or means that are greater than or equal to 64.90; under $H_1$, where we are assuming $\mu = 56$, $z$ for a mean of 55.10 = $(55.10-56)/2.5 = -0.36$, $z$ for a mean of 64.90 = $(64.90-56)/2.5 = 3.56$; for $z = -0.36$, area in smaller portion = .3594; for $z = 3.56$ area in smaller portion = .0002; power = .3594 + .0002 = .3596 or .36]

10. (a) **100** (2 pts) [required $\delta = 2.50$; $n = 2(2.50/0.5)^2 = 50$; 50 per group]
   
   (b) **15** (2 pts) [required $\delta = 2.50$; $n = 2(1-.70)(2.50/0.5)^2 = 15$]

   (c) **as the correlation gets smaller, power gets smaller** (1 pt)
   
   power will be the same as for an independent-samples $t$ test (1 pt)

11. **No** (1 pt)

   Even though there is substantial power to detect a large effect, the real effect size may be small and this study might not have adequate power to detect a small effect [or an effect that is arbitrarily small but not zero]. (1 pt)