Think about what needs to be measured in order to operationalize the research question in each of the questions listed above. Here are the ideas I can see that need to be measured:

1. How will we measure Spanish proficiency? What group will participants be in (we'll need at least two groups in order to show that it is the game playing that helps participants improve)?
2. What types of feedback will be provided and to how many groups? How will we measure grammatical accuracy and on what forms? How will we measure improvement in accuracy?
3. How do we measure how good people are at music? How do we measure how well they perceive Chinese tones?
4. How will we measure vocabulary knowledge? What group will participants be in (will there be a control group that does not try to learn any vocabulary at all)?
5. How will we measure phonological problems (will we target contrasts or specific sounds), and how many will we test?

2.1.1 Levels of Measurement of Variables

Answering these questions and thus operationalizing your research question will result in measuring variables. These variables may differ substantially from each other. For example, measuring Spanish speaking proficiency will probably be done through some kind of test where participants will receive a variety of scores. This type of variable is called a **continuous variable**. The values of a continuous variable consist of numbers that represent a measurement and involve a large range of possible answers. A continuous variable contrasts with a **categorical variable**, where the values of the variable basically do not have any inherent numerical value and consist of categories where one category is not rated any higher than any other. Categorical variables often have a very restricted range of possible values.

In the examples above, the number of groups that will be tested is a categorical variable, while scores on tests or measurements such as grammatical accuracy or perception of Chinese tones are continuous variables. Below I have listed the variables that will be needed for each of the operationalized variables in the five research questions my students asked, and a possible way of operationalizing the variable. Decide if each one is a continuous or categorical variable (I’ll give the answers following the list):

1. Spanish proficiency will be measured by an OPI. There will be three groups: those who play games, those who speak with a conversation partner, and those who do not do anything special.

2. There will be three types of feedback given to students on article usage: direct correction, indirect correction by circling, and no feedback. Grammatical accuracy will be measured by the number of accurate article usages out of all obligatory contexts. This will be measured on four different essays.

(Continued)
3 Facility with music will be measured through a perfect pitch test, with a maximum possible score of 20 points. Chinese tone perception will be measured through an identification test of four tones, with 30 items in total.
4 Vocabulary knowledge will be measured by having students put a check by the words they know and then totaling this number out of ten words.
   There will be three groups: those who learn multiword expressions, those who focus on single words, and those who are not specifically asked to memorize any words at all.
5 Five contrasts in English with ten items each will be played for the listeners, resulting in a test with a maximum possible score of 50 points.
   There are two groups of people with different first language (L1) backgrounds: Korean and Japanese.

Here are the answers: 1) continuous, categorical; 2) categorical, continuous; 3) continuous, continuous; 4) continuous, categorical; 5) continuous, categorical. It should be clear now that any type of grouping of participants will result in a categorical variable. In addition, whenever you give someone a test with a large number of items, this will result in a continuous variable.

I have started out with this broad division of measurements into continuous and categorical variables. The skill of identifying whether your variables are continuous or categorical is absolutely essential for deciding what statistical test to use, and in general the division between continuous and categorical is the most important division to know when it comes to performing statistical tests, as some tests can be used only when all variables are continuous, while other tests can be used only when there is at least one categorical and one continuous variable. However, many authors make a more detailed division of variable measurement into three or even four different types of measurement. For example, Mackey and Gass (2005) discuss three levels of measurement: nominal, ordinal, and interval scales. **Nominal scales** (which measure **categorical variables**) do not have any inherent numerical value. They are used to group people or data, and include measurements like gender, group membership in an experiment, first or second language background, or presence/absence of a linguistic feature. **Ordinal scales** are rank-ordered and do have some inherent numerical value. For example, we might have first-, second-, and third-year students of French. If we assigned the students a number based on their number of years of study, we would expect this to be reflected in their knowledge of French, with students with a higher number having more knowledge of French. We might not expect, however, that the difference between third-year students and second-year students would be the same as between second-year and first-year students. In other words, there is no guarantee that there are equal intervals of measurement between the participants. In an **interval scale** (which measures **continuous variables**), however, we would like to say that the difference between getting a 9 and a 10 on a 10-point test is the same as the difference between getting a 3 and a 4 on the test. In other words, the intervals are assumed to be equal. Some authors will also distinguish a **ratio scale** (Howell, 2002), which has all the properties of an interval scale plus the scale has a true zero point. I can’t think of any
Kondo-Brown (2006): Does accuracy in guessing the meaning of unfamiliar Japanese words differ depending on proficiency level?

| 1 | Proficiency level (divided participants into high and low proficiency on the basis of reading comprehension tests, with maximum possible score of 20 points). | continuous or categorical |
|   | Accuracy level measured on a series of different tests, but generally most of the tests had about 16 points or more. | continuous or categorical |

Boers, Eyckmans, Kappel, Stengers, and Demecheeleer (2006): Will explicit instruction in noticing formulaic sequences help one group speak more fluently than another group that does not get this instruction?

| 2 | Two experimental groups, one that was led to notice formulaic sequences and another that wasn’t. Fluency ratings done by judges who listened to taped conversations and rated them for fluency on a 20-point scale. | continuous or categorical |

Paavola, Kunnari, and Moilanen (2005): Does the number of times a mother changed her behavior because of a child’s action affect the number of phrases an infant understands?

| 3 | Number of times mother modifies behavior contingent on child’s action counted in two separate sessions of 20 minutes each (these are all added up together). Number of phrases the child knows based on the MacArthur Inventory parental report. | continuous or categorical |

Erdener and Burnham (2005): Does first language orthography affect how well participants can produce nonwords in a language that has an orthography similar or different to their L1?

| 4 | Two groups of participants with different orthographies (L1=Australian English or L1=Turkish). Two groups of target languages with different orthographies (L2=Irish English or L2=Spanish). Number of phoneme errors on 96 nonwords. | continuous or categorical |

Bialystok, Craik, Klein, and Viswanathan (2004): Do age and being bilingual affect the accuracy with which participants perform on the Simon test (a test that measures whether participants can ignore irrelevant information)?
some questions here where the answer is that there is no dependent or independent variable!

2.1.4 Application Activity: Practice in Identifying Variables

Write the independent and dependent variables in the tables. In some cases, there may be more than one independent variable, but there is never more than one dependent variable.

1 Wharton (2000): Do high-proficiency learners use more language learning strategies than low-proficiency learners?

<table>
<thead>
<tr>
<th>Independent:</th>
<th>Dependent:</th>
</tr>
</thead>
</table>

2 Munro, Derwing, and Morton (2006): Does L1 background influence the way that English learners rate the speech of other English learners for comprehensibility?

<table>
<thead>
<tr>
<th>Independent:</th>
<th>Dependent:</th>
</tr>
</thead>
</table>

3 Hirata (2004): Does pronunciation training with visual feedback help learners improve their perception of phonemically contrastive pairs of words more than training without visual feedback?

<table>
<thead>
<tr>
<th>Independent:</th>
<th>Dependent:</th>
</tr>
</thead>
</table>

4 Proctor, August, Carlo, and Snow (2006): What kind of relationship exists among an oral language measure and a reading comprehension measure in English for fourth-grade bilinguals?

<table>
<thead>
<tr>
<th>Independent:</th>
<th>Dependent:</th>
</tr>
</thead>
</table>

5 Bialystok, Craik, Klein, and Viswanathan (2004): Do age and being bilingual affect the accuracy with which participants perform on the Simon test (a test that measures whether participants can ignore irrelevant information)?

<table>
<thead>
<tr>
<th>Independent:</th>
<th>Dependent:</th>
</tr>
</thead>
</table>

6 Wartenburger et al. (2003): Do age of acquisition (AOA) and proficiency level affect what neural substrates are activated in the brain among different types of Italian–German bilinguals?

<table>
<thead>
<tr>
<th>Independent:</th>
<th>Dependent:</th>
</tr>
</thead>
</table>
7 Mizuno (1998) hypothesized that there would be a relationship between the amount of cultural exposure students had before they went overseas and the amount of cultural exposure they would obtain once they were living in the country.

8 Larson-Hall (2004): Do learners perform differently in their ability to perceive phonemic contrasts in nonwords in the second language depending on their proficiency level and on the sounds that are contrasted?

2.1.5 Summary of Variables

You should now be able to take a research question and decide what the variables are, whether they are categorical or continuous variables, and whether they are dependent or independent variables. Once you can take this step, you will be very far along on the path toward deciding what type of statistical test you'll need to analyze your data. For example, if you know that you have two continuous variables and that you are examining a relationship between the variables (and there is thus no independent or dependent variable) then you'll want to use the statistical technique of correlation. On the other hand, if you have a research question that has a dependent and an independent variable, and the dependent variable is continuous while the independent variable is categorical with only two groups, then you know you'll want to choose a t-test. We'll leave more of this type of fun until later (won't it make you feel smart to know what type of test you want?) and now delve into some abstract ideas about statistics that are also necessary for understanding how statistical testing works.

2.1.6 Fixed versus Random Effects (Advanced Topic)

In this book I will not generally distinguish between fixed and random effects. We will essentially treat all independent variables as fixed effects in regression and ANOVA models. A research model that contains both fixed and random effects is called a mixed-effects model, and this is not treated in this book. However, there may be times when it is important to understand the difference between fixed and random effects, so I will include that discussion here.

Fixed effects are those whose parameters are fixed and are the only ones we want to consider. Random effects are those effects where we want to generalize beyond the parameters that make up the variable. A “subject” term is clearly a random effect, because we want to generalize the results of our study beyond those particular indi-