Research Meeting Notes for March 24, 2006

Designing surveys for research
Surveys are one of the most frequently-used research instruments in education, and they are familiar to almost every college teacher and student. Despite their popularity, effectively designing and implementing surveys is a difficult and time-consuming enterprise. Wiersma (p. 178) identifies several phases and tasks associated with survey research, illustrated in the figure on the right.

While learning how to design, evaluate, and validate surveys is beyond the scope of this workshop, it is important to understand that the steps included in the picture actually represent the “tip of the iceberg” for survey design. Effective survey design also includes steps to validate the placement of items on the survey, identify correlations between survey items and established metrics, and determine if each survey item is phrased in the most appropriate manner possible.

Despite the intricacies and complexities associated with designing and validating surveys, the steps identified in the figure are a great place to start when conducting survey research, and following these steps can help ensure that surveys are developed in an effective and thoughtful manner.

Purpose of the survey
Although designing a survey may seem like an obvious choice for data collection in education, researchers need to evaluate whether the use of a survey represents the most effective and accurate means of data collection for research. Newly-developed surveys need to be validated in order for them to yield informative data, a process which can take hundreds of hours, multiple iterations, and continuous redesign of survey items. Often, other survey instruments may be available, or alternative forms of measurement may be more appropriate (such as conducting interviews, behavioral observations, and using direct measures of performance).

Determining if a survey is warranted
Trochim poses several questions that should be considered when determining if the use of a survey is appropriate. Some of these questions are as follows:

- **Are visual presentations possible?** When information cannot be effectively presented in a visual format, interviews and observational approaches may be more appropriate than surveys.
- **Are long response categories possible?** If so, then surveys should probably be avoided – open-response items (where respondents are encouraged to write in comments, suggestions, or other feedback) may not be effective if participants are tired, unmotivated, or unable to express themselves in writing.
- **Is privacy a feature?** Surveys may be seen as more anonymous than interviews and other research methods, and this perceived anonymity may increase the likelihood that respondents will be candid in their responses.
- **Is the method flexible?** Once designed, survey items are fixed and cannot be changed without influencing the validity of the survey instrument. In addition, the use of complex branching rules and instructions may not be well-understood by survey participants.
- **Is reading & writing needed?** Unless surveys are administered in person, with a third party present to explain items and interpret responses, participants must be able to read and write in order to comprehend and respond to the survey items.
• Can you judge quality of response? Because surveys are usually completed independently, it can be difficult to 
establish the validity of the responses, the truthfulness of the respondent, and the degree to which the respondent 
understood the question as intended.
• Are high response rates likely? Surveys can yield the lowest response rates among research methods, so data 
collection measures need to be taken to ensure that a sufficient number of responses will be received.
• Does it give access to dispersed samples? Surveys can provide access to diverse groups of participants located across 
a wide range of environments and settings.
• Does respondent have time to formulate answers? Respondents can answer survey questions on their own schedule, 
possibly leaving questions to be answered after some time has passed.
• Is there quick turnaround? Survey respondents may not be able (or motivated) to answer in a timely manner, let 
alone be aware that a rapid turnaround is required for the survey.

Constructing survey items
Wiersma (p. 179) provides several guidelines to follow when constructing survey items. Perhaps the most important 
guideline to keep in mind when developing survey items is the Law of Parsimony: “Keep things as simple as possible 
to obtain the necessary data.”

<table>
<thead>
<tr>
<th>Guideline (adapted from Wiersma)</th>
<th>Recommendations / Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. All items (other than demographic and background questions) should directly relate to the research problem, question, or hypothesis under investigation.</td>
<td>Avoid questions that seek additional data beyond the scope of the study – they increase the length and dilute the focus of the survey. Interviews, focus groups, journals, and observational methods may be more appropriate for gathering additional information.</td>
</tr>
<tr>
<td>2. Develop items that are clear, concise, unambiguous, and understandable by the targeted survey respondents.</td>
<td>Avoid vague terms that may be interpreted in different ways by different respondents; avoid jargon and technical terms that may not be well understood by all respondents.</td>
</tr>
<tr>
<td>3. Ensure that each survey item deals with only one subject or concept.</td>
<td>The following question should be broken into 2 separate questions: “Do you agree that students are best served by multi-modal instruction methods and open-seating arrangements?”</td>
</tr>
<tr>
<td>4. Avoid the use of leading questions that have implied assumptions, anticipated outcomes, or suggest a preferred response.</td>
<td>Wiersma’s example of a leading question is, “Are you in favor of relaxed discipline in the schools, even though such discipline undermines the moral development of youth?”</td>
</tr>
<tr>
<td>5. Avoid questions loaded with social or professional desirability such as those with responses which will make respondents disapprove of themselves.</td>
<td>Wiersma’s example of violating this guideline would be asking a teacher, “Do you have difficulty maintaining a good learning climate in your classroom?”</td>
</tr>
<tr>
<td>6. Avoid questions that demand personal or delicate information such as income, age, or engagement in illicit activities.</td>
<td>These questions can sometimes be asked using “response bands” such as “21-30 years old; 31-40 years old, 41-50 years old.” When asking personal or delicate questions, the truthfulness of the response must be considered.</td>
</tr>
<tr>
<td>7. Only request information that the respondent is able to provide.</td>
<td>Some respondents may not know the answers to the questions posed (e.g. “What is your credit score?”), so providing a “Don’t know” or “Not applicable” response can be informative. Also, consider using a “Confidence question” to ask respondents their level of confidence in the answer: “How confident are you that you know your credit score?”</td>
</tr>
<tr>
<td>8. Ensure that the reading level of the items is appropriate for the respondents.</td>
<td>Having representative people fill out and talk-through the survey items during a pilot study can help you ascertain if the reading level is appropriate for respondents.</td>
</tr>
<tr>
<td>9. Use “soft” words should be used instead of “hard” words (e.g. using “correction” or “corrective action” instead of “punishment.”)</td>
<td>Wiersma suggests this guideline as part of #8 above, but I don’t necessarily agree with it. While “soft” words may encourage respondents to be more forthcoming, these words can misrepresent the concept in question. Always try to provide an operational definition with examples to ensure that respondents understand the concept in question.</td>
</tr>
<tr>
<td>10. Favor short items over longer items, and simple items to complex items.</td>
<td>Wiersma suggests that using two or more shorter questions is better than using a single detailed and complex item. This is especially important if respondents are not well-versed in the terminology being used, or if they have memory or comprehension issues.</td>
</tr>
</tbody>
</table>
### Guideline (adapted from Wiersma) | Recommendations / Examples
---|---
11. Ask for specific numbers instead of averages. | People are poor estimators – they are better able to recall specific frequencies and instances than estimating averages. Instead of asking, “On average, how many presentations do you give per month?” ask “How many presentations have you given in the past month?”
12. Ensure that item response options are exhaustive and mutually exclusive. | Pilot test all items to ensure that responses make sense and cover all relevant possibilities from the respondent’s point of view. For example, don’t ask “How many years of college have you completed?” and then provide only positive integer responses (e.g. 1, 2, 3), as some respondents may need to respond with 0.
13. Provide neutral, middle-of-the-road, or “undecided” responses when it makes sense to do so, and when such answers would be meaningful. | Wiersma suggests this as part of guideline #12 above, and applies it to all items. I don’t agree that neutral responses are appropriate for all items – respondents may opt for these responses to avoid undesirable responses or save mental energy instead of thinking about subtle differences.
14. Phrase questions positively and avoid questions with negative and double-negative expressions. | Although negative questions are sometime necessary, most questions should be worded positively to help the respondent understand the question. Don’t use phrases like “Which reading techniques do you avoid?” and “When would you not avoid these techniques?”

### Validating responses to survey items
Survey responses are only as good as the validity of the data from respondents, who may provide erroneous, inaccurate, or deliberately misleading information for a number of reasons. While there are no guaranteed methods for ensuring the accuracy of the responses, several methods can be used to improve your ability to identify invalid survey responses. Wiersma cites the following four techniques developed by Wrona, Sanborn, and Constantine (1992), who used these techniques to help assess the validity of responses to a survey about drug use.

- **Impossible response:** Item responses can be used to assess validity by including responses that are not possible (such as a course that doesn’t exist, or a class time that doesn’t match that of an actual course). Unfortunately, including impossible responses in a survey item can increase the number of responses available (contributing to a more confusing item) or reduce the number of valid items that can be included. This technique is often used in market research, when participants are asked about which brands that they’ve used before.

- **Implausible frequency and extent of use:** Some responses that deal with frequencies and durations can be evaluated for plausibility, and items that seem implausible can be flagged for further scrutiny. For example, while it’s possible that a person could study for 18 hours per day, every day, during a semester, it’s more likely that the respondent is exaggerating or misreporting the true frequency.

- **Unlikely response combinations:** Patterns of responses between items can often be evaluated to determine inter-item consistency. For example, a respondent who states that she studied for 12 hours on Monday in one question, but also slept 8 hours on Monday in another question, and worked 12 hours on Monday in yet a third question, is providing an unlikely response combination. This participant may be providing incorrect information, or perhaps had trouble comprehending the questions.

- **Honesty item:** The most straightforward approach to assessing validity is to ask respondents to indicate the level of honesty they exhibited in the survey. Honesty items can encompass the entire survey, or specific sections, such as “Indicate how honest you were when answering the questions on drug use.” Because honesty may be interpreted in different degrees, it can be informative to ask honesty items on a scale (e.g. from “Completely honest,” “Somewhat honest,” “Neutral,” “Somewhat dishonest,” “Completely dishonest”) instead of a binary “yes” or “no.”

When evidence of misrepresented, incorrect, or dishonest answers is discovered, the survey should be “flagged” (coded as having potentially invalid responses) and then evaluated in more detail for additional evidence of invalid information. Responses from these surveys can still be included in the main dataset to see if the responses change the results of the study. If the responses do influence the study results, these surveys should be set aside and analyzed separately. These surveys may be thrown out as outliers, or included in the analysis, but reported separately if they change the results of the study. (Regardless of the approach used to handle these surveys, the analysis procedures and results should be clearly explained in any write-up of the study.)