

CHAPTER 4

Item-Writing Guidelines

The words used in a question will always influence how study participants answer that question.

—PETERSON (2000, p. 46)

Introduction

Peterson's statement reminds us of the critical choices we make in writing survey items. Given that the words in a survey item will influence the response of a study participant, we want to take care to write items in a way that the responses will allow us to make valid inferences (e.g., decisions, predictions) about the study participants. In Chapter 4, we present guidelines that will help survey developers to write items that are clear to the respondents and that will assess the construct of interest and its domains. Dimensions to guide item writing include (1) relevance to the survey purpose and construct, (2) audience, (3) language, (4) item structure, and (5) conventions.

Addressing Relevance

Relevance is the degree to which an item aligns with the construct and domains that are the focus of the survey scale. The following guidelines will assist you in writing items that support relevance and limit the introduction of error into your survey results.

The Construct of Interest

Relevance requires that an item measures the concept or construct that is the focus of the survey scale (see Table 4.1). Before you begin writing items, you should make explicit the construct and its domains that you want to measure with the survey scale. Specifying the construct and its domains supports the item writer in determining whether an item is related to the survey purpose (DeVellis, 2012).

Return to the conceptual framework in Figure 3.3 and the depiction of the constructs related to student transition/retention status (i.e., passing or failing). In this instance, relevant construct domains include social engagement and academic engagement. This listing reminds us of the construct domains that we should include in the survey and prevents us from including irrelevant items.

Review the item in Figure 4.1 to consider whether it should be included in a student engagement scale. If the domain of interest in Figure 4.1 is student *social* engagement in the high school setting, a question arises about the appropriateness of the item “As a ninth-grade student I know my way around the school.” In a social engagement scale, the purpose is to assess the student’s connection to other students and his or her teachers. Thus, this item might be deleted from the pool of survey items related to the domain of social engagement. Does the alternative item—“As a ninth-grade student I have many friends at school”—address the domain of social engagement? If so, then it is a candidate for the engagement scale.

Logically Related to the Construct

In beginning to write items, think creatively about the construct and its domains (DeVellis, 2012). Ask yourself, “What are alternative ways that I can word an item to measure the construct?” This process will be facilitated

TABLE 4.1. Guidelines for Addressing Relevance

- | |
|--|
| • Develop items that address the construct of interest. |
| • Write items that are logically related to the survey purpose. |
| • Use multiple items to tap into a construct, but avoid repetition of items. |
| • Avoid items that crossover to a related construct (e.g., social anxiety rather than test anxiety). |
| • Write items to be concrete and precise. |
| • Keep items and words within an item relevant. |
| • Keep items objective. |

		⊘ NOT THIS			
		Strongly disagree	Disagree	Agree	Strongly agree
7. As a ninth-grade student I . . .					
a. know my way around the school.		1	2	3	4
		✓ BUT THIS			
		Strongly disagree	Disagree	Agree	Strongly agree
7. As a ninth-grade student I . . .					
a. have many friends at school.		1	2	3	4

FIGURE 4.1. Revisions to items to support relevance.

if your writing is grounded in a review of the literature and examination of similar scales.

As part of the creative generation of items, return to the items that you draft and examine their relevance in gauging the construct of interest. This review will safeguard against the inclusion of items that are not logically related to the survey purpose. Reviewing the conceptual model helps to ensure that items are logically related to the construct domains. Consider, for example, writing items to assess factors related to successful student transition from eighth to ninth grade. At first glance, the statement “My teacher gives checklists or rubrics to us as we prepare to do projects” appears unrelated to the topic. However, if the conceptual model includes a teacher factor as a potential contributor to student transition (see Figure 3.3), then writing items for a subscale that measures teacher assessment practices is appropriate.

Multiple Items

In developing items for a scale, it is generally good practice to write multiple items that measure each domain of the construct that the survey scale measures. Writing multiple items for inclusion in a survey scale supports the researcher in better assessing the breadth of the construct and supporting the consistency of domain scores.

Multiple Items for Consistency

Typically, a score from a construct domain is composed of several items. For example, a classroom climate subscale might include multiple items that measure teacher instructional practices:

My teacher checks to see if we are understanding the classwork.

My teacher makes it easy for me to ask questions.

My teacher tells us what to study to prepare for a test.

(adapted from Johnson, Davis, Fisher, Johnson, & Somerindyke, 2000)

Using several items to form a score for the domain of classroom climate will support score consistency more than using a single item. Consistency of scores is related to the reliability of scores, a topic we explore in depth in Chapter 8.

The need for multiple items, however, is not addressed by repeating an item after minor changes to that item. Respondents might perceive repetitive items as irrelevant (Peterson, 2000). Consider the following item from a scale on classroom climate: “My teachers respect me.” Students might perceive as irrelevant a repetitious item that states, “I am respected by my teachers.” An item that taps into respect, but is not simply a repetition of another item, is “My teachers care about what I have to say.”

Multiple Items for Content Representativeness

Multiple items also are needed to measure a construct that is composed of several domains. Multiple items specific to each domain will ensure that the items are representative of the construct and its domains. In the Caregiver Survey (see Morgan et al., 2010: see Table 1.2), the scale has three domains: job satisfaction, supervisory relationships, and job environment. Thus, items had to be developed for all three domains. Examples of potential items that address a job environment domain include:

I am involved in challenging work.

I have a chance to gain new skills on the job.

Items for the supervisory relationship domain might include

My supervisor is open to new ideas.

My supervisor is available to answer questions when I need help with my clients.

My supervisor tells me when I am doing a good job.

My supervisor is responsive with problems that affect my job.

(adapted from Morgan et al., 2010)

Notice that the two domains of supervisory relationships and job environment represent two different, but related, expressions of a complex condition or phenomenon, that is, job satisfaction. Thus, we need

multiple items for purposes of consistency and for adequate assessment of the breadth of the construct.

Crossover

Relevance also requires avoiding items that cross over to a related construct. Such a crossover is evident in the item “As a ninth-grade student I care if I finish high school.” This statement appears to measure a generalized engagement (or disengagement) of the student rather than academic or social engagement.

Concrete and Precise

Writing items that are concrete and precise will avoid ambiguity. Items gauging teachers’ attitudes about inclusion of students with special needs in the classroom illustrate the need for items to be unambiguous. Contrast the following statements related to inclusion of students with disabilities in a general education classroom:

Inclusion is good for all children.

Inclusion is good for most children.

(Stanley, n.d.)

Does a high score on inclusion being good for *all* children translate to a high score for the construct of teacher’s attitude toward inclusion? Then does a high score on inclusion being good for *most* children translate to a high score for the construct of teacher’s attitude toward inclusion? Does a teacher indicating that inclusion is good for *most* children translate to a low score for the construct? Thus, the items are ambiguous in terms of a teacher’s attitude toward inclusion.

Item and Word Relevance

Survey scale items must be relevant, and the words that constitute an item must be relevant. Consider using the item “As a ninth-grade student I come to school only because my friends are here.” The intent of the item appears appropriate—to gauge the importance of a student’s friends in motivating him or her to come to school. However, the term “only” muddies the meaning. If the student strongly agrees, then is that response—“I strongly agree that I come to school *only* because my friends are here”—positive in terms of social engagement in a school setting? If the item is revised to say that “I come to school to see my friends,” then the focus of the item appears to more closely gauge the construct of social engagement in the school setting.

Objectivity

Relevance also requires the survey developer to maintain objectivity in items. Nonobjective items suggest the correct answer; such leading items discredit the researcher. A nonobjective item uses emotionally laden language that will introduce bias and interfere with the interpretation of respondents' answers. One will not know if responses reflect study participants' reaction to the biased language or to the content of the items. The following item is from the Research Committee of the Sierra Club (n.d.):

Our nation is still blessed with millions of acres of public lands, including roadless wilderness areas, forests and range lands. Land developers, loggers, and mining and oil companies want to increase their operations on these public lands. *Do you think these remaining pristine areas of your public lands should be protected from such exploitation?*

Yes No No Opinion

The emotionally charged language in this item makes clear the answer desired by the surveyors. The bias in the item is likely to call into question the credibility of the research committee conducting the survey.

Addressing Audience

A major consideration in writing items is that of audience (see Table 4.2). In considering audience, we ask, "What are the characteristics of the survey respondents?"

Cognitive Skills and Communication

As you write items, you should take into account the cognitive skills and communication capabilities of the survey respondents. For example, will respondents be children or adults? Is the language that you use in your items

TABLE 4.2. Guidelines for Audience

- Consider the cognitive skills and communication capabilities of the respondents and their ability to understand an item.
- Determine if respondents will have sufficient information to answer the items.
- Consider if the study participants will be able to recall the information (e.g., behaviors, activities).
- Represent the diversity of respondents (i.e., multicultural, nonbias).

appropriate for the age group? For high school students, you might use the following statement, “Participation in the *Discovery Workshop* improved my written communication skills.” However, for elementary students you might say, “Being in the *Discovery Workshop* improved my writing.” The reading level is demanding for both items; however, the reading level for the latter statement for the elementary students is less demanding than the item written for high school students.

Sufficient Information

Determining whether respondents will have sufficient information to answer items will contribute to the development of appropriate items. Ask yourself, “Will respondents have sufficient knowledge to answer this item?” This guideline addresses the ability of the respondent to provide the desired information. We distinguish this guideline from the respondent’s understanding of the topic (i.e., knowledge), which is addressed later in this chapter. Figure 4.2 provides an illustration of an item that asks students how frequently their teacher meets with students who are having difficulty with their work. In deciding whether this item should be used in a student survey, you should ask yourself whether students are likely to know if a teacher assists *other* students before or after school. Do not expect accurate answers when asking respondents what others do or feel (Nardi, 2003).

A second issue is whether the students are likely to know how frequently a teacher meets with students who need assistance. Although a student might know that her teacher has a policy of assisting students during lunch, the student is unlikely to know the frequency with which other students receive assistance from the teacher during lunch.

Recall of Information

Respondents’ ability to recall information about activities and behaviors should also be considered. In gauging respondents’ ability to recall information, use a time period related to the importance of the item topic. Peterson

4. My English teacher . . .	Rarely	Sometimes	Often	Frequently
f. meets during lunch with students who are having difficulty with their work.	1	2	3	4

FIGURE 4.2. Example of an item that asks for information that is unavailable to the respondent.

(2000) warned against asking respondents about practices over a long time period and recommended focusing on a short time period for which the respondent is likely to recall events or daily practices. Important information, such as major life events, can be recalled over a longer time period, whereas commonplace events might be recalled for only the previous few weeks or months (Fink, 2003; Fowler, 1995).

Representation of Diversity

Attention to diversity is another aspect of audience. In developing items, the writer should consider the diversity of the survey respondents and ensure the various groups would perceive the survey as relevant. For example, retirees might consider irrelevant any items that ask about behaviors in the office setting. The relevance of the language used in the items should also be reviewed by members of the various groups that will be completing the survey scale. This review can be achieved by conducting focus groups with participants similar to those who will complete the **operational** survey.

The issue of audience cannot be separated from language, conventions, and item structure. Thus, as you develop items, you should revisit the issue of audience to inform decisions related to these qualities of surveys.

Addressing Language

In the writing of an item, attention should be paid to the language in order to craft items that the study participants will understand. Contrast presenting parents with the stimulus “I visit the school for educational functions” versus the statement “I attend school events.” Some parents might be confused by the term “functions” but understand the word “events.” This example shows the importance of attending to word choice and the demands of language in writing an item.

Language Understood by Respondents

The first guideline in Table 4.3 advises using language that will be understood by the study participants. The item in Figure 4.3 uses straightforward language—*makes it easy for me to ask questions* rather than *facilitates asking questions*. In using the term *me*, the item developer expresses the item in the context of the student’s experience.

Respondents must be familiar with the vocabulary used in the items. Study participants might respond to items with unfamiliar vocabulary; however, their responses might be inconsistent owing to their partial understanding of the item. Recall that the intent of the researcher is to make inferences/draw conclusions about the participants’ attitudes, knowledge,

TABLE 4.3. Guidelines for Addressing Language (Vocabulary)

• Use words with meanings understood by respondents.
• Use language (i.e., terms, concepts) that is current.
• Avoid words with multiple meanings.
• Attend to word choices.
• Avoid abstractions.
• Write items at middle school level reading demands for the general public.

or behaviors. A researcher might arrive at inappropriate conclusions if the language demands of survey items result in responses counter to the intent of study participants. For example, in a school climate survey, some parents might correctly interpret the phrase “students retained” to mean the percentage of students who were not promoted to the next grade level, whereas other parents might interpret the phrase as providing information about the graduation rate. Respondents’ interpretation of a statistic, for example, 6.2%, is likely to be very different if parents think the statistic reflects the percentage of students who must repeat a grade versus the percentage of students who graduate. Although parents who initially interpret the information as being related to graduation will likely self-correct, in developing a survey scale we want to write items that are clear in their intent.

Items with numerous technical terms might result in respondents appearing to have little knowledge about a topic; however, if the items had been presented in lay terms, then the survey respondents could have demonstrated their understanding. In Figure 4.4, the item could be edited to delete the term “transition” and make the item more understandable to students.

	⊘ NOT THIS			
3. My teacher . . .	Rarely	Sometimes	Often	Frequently
e. facilitates asking questions.	1	2	3	4
	✓ BUT THIS			
3. My teacher . . .	Rarely	Sometimes	Often	Frequently
e. makes it easy for me to ask questions.	1	2	3	4

FIGURE 4.3. Example of an item that uses the language of the respondent.

Based on your opinion about how well you are doing in school, indicate how well the following aspects of the Quest Center prepared you for middle school.

⊘ NOT THIS

a. The math round tables at the Quest Center prepared me for transition to middle school mathematics.	1	2	3	4
	Strongly disagree	Disagree	Agree	Strongly agree

✓ BUT THIS

a. The math round tables at the Quest Center prepared me for mathematics in middle school.	1	2	3	4
	Strongly disagree	Disagree	Agree	Strongly agree

FIGURE 4.4. Example of an item with demanding vocabulary.

Current Language

Another consideration in writing items is to use language that is current. For example, in special education, one no longer refers to a learning-disabled child; rather, one speaks of a child with a learning disability (University of Kansas Research and Training Center on Independent Living, 2008). Instead of using the term “homosexual,” one should use “gay men” or “lesbians” (APA, 2010). Notice that the language here also overlaps with the issue of diversity discussed previously.

Multiple-Meaning Words

Words with multiple meanings have the potential to create confusion, so they should be avoided. The phrase “school environment information,” for example, has multiple meanings. Does the researcher want information about the physical environment, such as air quality, structural integrity, or handicap accessibility? Or do researchers want information about teacher and student morale, parental involvement, and school safety? If multiple-meaning words cannot be avoided, then the context of the item should clearly establish the meaning of the word within the item.

Choices about Words

To improve respondent understanding of survey items, one should attend to word choices and select wording that is meaningful to study participants. So, when writing, avoid highly technical words, infrequently used words, slang, colloquialisms, jargon, and abbreviations. You should also

avoid words with multiple syllables. As with any guideline, there will be exceptions to the rule. For example, the technical language associated with a field, such as public health, might appear to be jargon for those outside of the field but convey the appropriate meaning to public health educators. For instance, although the term “body mass index” and its related acronym, BMI, are commonplace in the public health field, for those outside the field the language might be unfamiliar.

Abstractions

In selecting wording, you should avoid abstractions. For example, do not ask about behavior in terms of “average.” Asking a respondent the number of books that he or she reads on average each month is a cognitively demanding task. It is better to ask the typical number of books a respondent reads each month. In Figure 4.5, the phrase “climate for implementation” is an abstraction. What is a climate for implementation? Does this have to do with principal support? Is it in reference to professional training having been completed to prepare teachers for implementation of a magnet program? Is the term addressing parent support of a magnet? If it is the latter, then the item could be written as “Parents are supportive of the magnet program.”

Reading Demands

For the general public, the reading demands of a survey should be at the fifth- to eighth-grade level (Barnette, 2010; Peterson, 2000). The spelling and grammar tool in Microsoft Word provides an estimate of the reading level for a document (Microsoft Corporation, 2010). To obtain an estimate for an item or a document, highlight the text for which you want to estimate a reading level, click the Review tab, and click the Spelling & Grammar option. A pop-up window states that Word has completed checking the selection and asks whether you want it to check the rest of the document. When you select No, another pop-up with the readability indices appears.¹

Two useful indices of reading levels are the Flesch–Kincaid Grade Level and the Flesch Reading Ease (Microsoft Corporation, 2010). In Figure 4.5, the top item on a positive school climate has a Flesch–Kincaid Grade Level index of 9.4, which means a ninth-grade student should understand the text. In contrast, the bottom item on supportive parents in Figure 4.5 has a Flesch–Kincaid Grade Level index of 7.3 (seventh grade, third month), indicating the revised item is slightly easier to read.

¹In Word 2013, if a pop-up window does not appear with the readability statistics, then go to REVIEW → LANGUAGE → LANGUAGE PREFERENCES → PROOFING and check the “show readability statistics” option.

⊘ NOT THIS						
	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
23. The climate for implementation of the magnet program is positive in my school.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
✓ BUT THIS						
	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
23. Parents are supportive of the magnet program.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE 4.5. Example of an item with an abstraction.

The Flesch Reading Ease has a 100-point scale. As the score increases, the document is easier to read. With a score of 54.7, the item about parent support in Figure 4.5 is fairly easy to read. The scores for both readability indices are based on the average number of syllables per word and the number of words per sentence (Microsoft Corporation, 2010). Thus, you can control the reading level required by your assessments by avoiding words with multiple syllables and by attending to the length of the sentences.

Addressing Item Structure

The complexity of an item can be controlled by following the guidelines in Table 4.4. Items should be as brief as possible and still clearly communicate to the survey respondent what the researcher is asking.

Brevity

Peterson (2000) recommended keeping items brief. To achieve brevity, he recommended limiting an item to 20 words or less and using no more than three commas. In a survey, each word is associated with a cost, such as the direct cost of web development and the indirect cost of a respondent's time. Adherence to the guideline of brevity helps to control these costs. Recall also that reading level is, in part, a function of sentence length and brevity will reduce reading demands.

TABLE 4.4. Guidelines for Addressing Item Structure

- | |
|--|
| • Write items to be brief. |
| • Write complete statements. |
| • Present a single idea in an item. |
| • Use positive wording instead of negative phrasing. |
| • Begin items with qualifying clauses because respondents begin to formulate a response before the item is complete. |
| • Reduce the reading load by eliminating repetition in the phrasing of items. |

Complete Sentences

To keep items short and simple, you might be tempted to use phrases for some items. For example, in the demographics section of a survey you might be tempted to use the following:

Year of birth?

Survey developers, however, warn against such a shortcut. Using phrases instead of complete statements likely increases the cognitive demands for the respondent. In essence, the survey respondent must fill in the gaps and mentally “write” the item to read, “What is your year of birth?” Thus, to keep the demands of the survey low for the respondents, items should be written as complete statements, not as phrases.

Single Idea

The complexity of an item is reduced, and the interpretability increases, when a statement or question focuses only on one idea. As shown in Figure 4.6, the survey respondents are asked whether their English teachers write comments about the strengths and weaknesses of the students’ graded work. This item fails to focus on one idea. Is a student to respond in regards to his or her teacher’s written comments about weaknesses or to the teacher’s comments about strengths? Some teachers mark weaknesses; so perhaps the student should circle *Often* or *Almost all the time*. But what if the teacher rarely writes comments about students’ strengths? If the items were written to focus on one idea, then the task is straightforward. One way to guard against items with more than one focus is to watch for use of the word “and” in an item. Often the term “and” is used to join two ideas in an item and serves as a cue that the ideas should be split into two items.

⊘ NOT THIS				
4. My English teacher . . .	Rarely	Sometimes	Often	Almost all the time
f. writes comments about the strengths and weaknesses of my graded work.	1	2	3	4
✓ BUT THIS				
4. My English teacher . . .	Rarely	Sometimes	Often	Almost all the time
f. writes comments about the <i>strengths</i> of my graded work.	1	2	3	4
g. writes comments about the <i>weaknesses</i> of my graded work.	1	2	3	4

FIGURE 4.6. Example of an item that focuses on a single idea.

An item that gauges teachers' attitudes toward inclusion provides another illustration of the need for an item to have a single focus. Classroom teachers were asked to respond to the following item, "Working with the consultant teacher takes valuable time away from planning and student time." If a teacher strongly agrees with the statement, then is the problem with planning time, student time, or both?

The presentation of a single idea is critical in the use of true–false items. Examine the following item. This item asks about alcohol *and* smoking.

5. Women who smoke and drink more than one alcoholic beverage a day increase their risk for breast cancer.

(adapted from University of Rochester Medical Center, 2014)

If a survey respondent answers this item correctly, is it because he knew that both smoking and alcohol were risks associated with breast cancer? Or did the participant answer it correctly because he knew smoking increased the risk but he was less certain about alcohol consumption?

Positive Wording

Using positive wording instead of negative phrasing also reduces the complexity of an item. For example, a survey respondent must perform mental

gymnastics to answer the item “The budget for teacher resources at the Quest Center is *not* adequate.” A respondent might ask herself, “If I mark strongly disagree, then I think the budget is adequate, right? If I mark strongly agree, then the budget is not adequate?” That the item requires the respondent to decode its meaning is problematic. Faced with multiple items that are cognitively demanding, the respondent might not complete the survey.

Qualifying Phrases

Qualifying clauses should occur at the beginning of a survey item because respondents begin to formulate a response before the item is complete (Peterson, 2000). So, if an item states, “I really feel that I am a member of a team within the FAME magnet program,” then a teacher might formulate his response based on his sense of belonging to the school in general (i.e., “I really feel that I am a member of a team”). To gauge the teacher’s sense of belonging to the magnet program, the item should be phrased, “Within the FAME magnet program, I really feel that I am a member of a team.”

Eliminating Repetitive Phrasing

Items within a scale often begin with similar phrasing. In Figure 4.7, each item begins with the phrasing “Consultation on. . . .” The reading demands of the item can be reduced by creating a stem with the repeated words. Note the ease in determining the focus of each item when the repetitive material is removed.

Addressing Conventions

Attending to conventions helps to create a professional look for a survey. A survey that fails to follow conventions and is riddled with typographical errors sends the wrong message to survey respondents about their value. The guidelines offered in this section provide reminders about creating a survey that conveys to survey respondents their importance to a study.

Language Conventions

In writing items, Peterson (2000) advised following standard language conventions, such as avoiding double negatives in items. As shown in Table 4.5, the items in the survey scale should be reviewed to ensure the items follow standard language conventions, including spelling, punctuation, and

⊘ NOT THIS				
<i>Satisfaction with Communication Center Services for Students</i>	Strongly dissatisfied	Dissatisfied	Satisfied	Strongly satisfied
Consultation on resume writing	1	2	3	4
Consultation on academic assignments	1	2	3	4
Consultation on business writing	1	2	3	4
Consultation on thesis/dissertation	1	2	3	4
✓ BUT THIS				
<i>Satisfaction with Communication Center Services for Students</i>	Strongly dissatisfied	Dissatisfied	Satisfied	Strongly satisfied
Consultation on . . .				
academic assignments	1	2	3	4
business writing	1	2	3	4
resume writing	1	2	3	4
thesis/dissertation	1	2	3	4

FIGURE 4.7. Example of reducing the reading load of an item.

grammar. Attention to conventions will reduce the likelihood that errors will interfere with survey participants' understanding and their responses.

Typographical Errors

As the drafting of items comes to a close, you need to review the items for typographical errors. Microsoft Word uses red underlining to show potential spelling errors or green underlining to indicate possible grammatical

TABLE 4.5. Guidelines for Addressing Language Conventions

- Follow standard language conventions.
- Review for typographical errors.

issues. However, in developing items and attending to all the guidelines that we have discussed, it is easy to overlook markings and not correct typographical errors. That is, sometimes in our haste to complete the development of a survey, we forget to page through the document for errors. Also, sometimes incorrectly used terms are not flagged by a word-processing application because they are correctly spelled, but they are not the intended words. For example, if I misspelled “words” as “works,” a word-processing program might not flag it because it is a correctly spelled word.

A survey with typographical errors might create confusion for the respondents. Also, typographical errors might raise questions about the researcher’s credibility. Surveys that have been edited for conventions and typographical errors express a sense of professionalism and convey to respondents the importance of their role in the research effort.

Guidelines Specific to Item Type

In the following section, we present guidelines specific to item type: knowledge, behavior, and demographics.

Guidelines Specific to Knowledge-Based Items

In Chapter 1, we raised the possibility of using true–false item formats to gauge survey respondents’ knowledge of a topic. The item-writing guidelines in the previous sections of this chapter apply to developing knowledge-based, true–false items. That is, in writing true–false items, one must be sure that the items align with the construct that is the focus of the survey scale. The audience must be considered; the language written must be clear and at the appropriate reading level for the survey respondents; and items written need to be brief and to clearly communicate the intent of the items and the conventions of language followed. Although there is considerable overlap in guidelines for item writing, the true–false format has some issues specific to that type of item. Guidelines specific to writing true–false items are listed in Table 4.6.

TABLE 4.6. Guidelines Specific to Knowledge Items

- | |
|---|
| <ul style="list-style-type: none"> • Do not mix partly true with partly false statements. • Avoid a disproportionate number of correct true and correct false responses. • Avoid specific determiners (e.g., “all,” “none”). • Consider the use of variants of the true–false form. |
|---|

Partly True–Partly False

In the use of true–false items, do not mix partially true with partially false statements. If we want clarity in an item, then the statement must be either true or false, not both. Consider the following statement:

3. A healthy diet includes plenty of whole grains, fruits, vegetables, and saturated fat.

(adapted from University of Arkansas for Medical Sciences, 2014)

A healthy diet does include plenty of whole grains, fruits, and vegetables, so the statement appears true. However, plenty of saturated fat is not part of a healthy diet, so is the statement false? We do not want to require survey respondents to internally debate the intent of the true–false statement. Better to use two items:

- A healthy diet includes plenty of whole grains, fruits, and vegetables.
- Saturated fat is part of a balanced diet.

Balancing True–False Statements

Should you use mainly true statements, more false than true items, or equal amounts of both? Early researchers advised using more false items because when respondents are uncertain of an answer, they are likely to guess the answer is true (Cronbach, 1946, 1950; Gustav, 1963). Ebel (1965) suggested that 60% of items should be false.

Specific Determiners

Specific determiners, such as “all,” “always,” and “never,” cue the respondent that a statement is likely false. In contrast, modifiers such as “generally,” “sometimes,” and “many” provide a clue that a statement is true. Hopkins (1998) suggested that balancing the number of true and false items that use specific determiners can decrease the influence of such modifiers on respondents’ answers.

Variants of the True–False Form

A variation of the true–false format involves using an item format with alternate choices. Figure 4.8 shows two items that provide alternate choices for respondents to select. Each item presents a brief scenario that describes classroom assessment practices that are ethical or unethical (Johnson, Green, Kim, & Pope, 2008).

<p>1. A teacher assesses student knowledge by using many types of assessments: multiple-choice tests, essays, projects, portfolios.</p> <p><input type="radio"/> Ethical</p> <p><input type="radio"/> Unethical</p>
<p>16. A teacher allows a student with a learning disability in language arts to use a tape-recorder when the student answers the essay questions on social studies tests.</p> <p><input type="radio"/> Ethical</p> <p><input type="radio"/> Unethical</p>

FIGURE 4.8. Alternate choice items for assessing educators' understanding of ethical classroom assessment practices.

Study participants (preservice teachers, teachers, and administrators) marked whether each assessment practice was ethical or unethical. Based on the literature on assessment, Johnson et al. (2008) developed an answer guide and scored participants' responses as correct or incorrect. These scores provided information about the survey participants' understanding of ethical practices. To the degree that respondents' selection of myth/fact and ethical/unethical scenarios are similar to the selection of true and false items, more myth items and more unethical items appear appropriate (see the earlier section "Balancing True-False Statements").

Related to the true-false format is the checklist format. The items shown in Figure 4.9 can serve to gauge teachers' understanding of the types of assessments that are appropriate for assessing students' progress in reading. The scoring of teachers' responses would be guided by the researcher's theoretical framework of reading. For example, whole language focuses on how students make sense of the text as they read (Taylor, 2007). Teaching and assessment within this framework is characterized by the principle of authenticity of texts, tasks, and tests and integration of reading, writing, speaking, and listening (Pearson, 2004). Within this framework, if a teacher checks portfolios as one appropriate method for monitoring student reading progress, then the researcher can count the response correct if portfolios provide information about how students make meaning of text as they read. In contrast, if a teacher checks end-of-chapter core reading tests (i.e., multiple-choice tests), then the researcher might consider the response incorrect because such tests often rely on reading passages that are written for the sole purpose of testing reading skills and do not engage students in authentic reading experiences. Thus, the researcher would review each classroom assessment, use the theoretical framework to determine if each

Please indicate which of the following classroom assessments are appropriate for monitoring students' reading progress. **Please select all that apply.**

- Anecdotal notes
- Conferencing with students
- Core reading tests (supplied by publisher)
- Miscue analysis
- Observation
- Running records
- Spelling tests
- Student portfolios
- Teacher-made tests (e.g., multiple choice, short answer, matching)
- Vocabulary tests
- Worksheets
- Writing samples
- Other (Please specify in box below)

FIGURE 4.9. Checklist to gauge teachers' understanding of appropriate assessment practices.

assessment practice was appropriate, develop a key for scoring teachers' responses, and score and tally each teacher's correct responses.

Assessing Knowledge versus Perceptions about Knowledge

The intent of the previous examples is to test respondents to measure their understanding of important knowledge. In some instances, the researcher might develop a survey scale in which respondents self-report *perceptions* about their knowledge level. The items in Figure 4.10 do not actually "test" a respondent to measure his or her knowledge about a topic; rather, the items ask for the respondent's perceptions about his or her understanding of technology.

Guidelines for Items about Behaviors

A survey scale also might address behaviors (e.g., walking, jogging, golf, basketball; Peterson, 2000). When measuring behaviors, items are often accompanied by a frequency response scale, which will be discussed in more detail in Chapter 5. If one wants to measure parent involvement in their

I know how to . . .		Yes	No
a.	turn the desktop on and off.	<input type="radio"/>	<input type="radio"/>
b.	log onto the network.	<input type="radio"/>	<input type="radio"/>
c.	conduct research on the Internet.	<input type="radio"/>	<input type="radio"/>
d.	use Microsoft Word.	<input type="radio"/>	<input type="radio"/>
e.	use Microsoft Publisher.	<input type="radio"/>	<input type="radio"/>
f.	use Microsoft PowerPoint.	<input type="radio"/>	<input type="radio"/>
g.	use Microsoft Excel.	<input type="radio"/>	<input type="radio"/>
h.	use CPS clickers.	<input type="radio"/>	<input type="radio"/>
i.	use SMARTBoard (write on it, erase, and move objects).	<input type="radio"/>	<input type="radio"/>

FIGURE 4.10. Example of an item that assesses student *perceptions* about knowledge.

child's school, then the survey scale might include the item in Figure 4.11. Note that the response scale is a frequency scale ("Rarely"–"Frequently") and not a satisfaction scale (i.e., "Dissatisfied"–"Satisfied").

Parent involvement could also be measured by using a checklist of behaviors as shown in Figure 4.12. Contrast the Likert-style item in Figure 4.11 with the checklist-formatted item in Figure 4.12. The major difference is that the checklist captures the different types of activities parents might engage in but not the frequency of the behaviors. The list of behaviors could be developed through reviewing the literature and conducting a focus group with some parents, perhaps the members of a school improvement committee or the Parent, Teacher, Student Association.

To ensure that all respondents consider the same timeframe, you should specify the time period for which they report their activities. Recall that earlier we discussed the Peterson's recommendation (2000) to avoid asking respondents about practices over a long time period if respondents will be asked to recall events or daily practices. However, Fink (1995) warned to avoid too short a time period because less frequently occurring events may

Academic Year 2017	Rarely	Sometimes	Often	Frequently
I attended school events.	1	2	3	4

FIGURE 4.11. Example of an item to measure parents' involvement in their children's school.

22. Please check all the volunteer activities in which you participated at the Center for the Arts during the 2016–2017 school year.

- act as a homeroom parent
- accompany students on field trips
- listen to students read
- participate as a classroom volunteer
- read to students
- serve on the School Improvement Committee
- tutor students

FIGURE 4.12. Example of a checklist item to measure parents' involvement in their children's school.

be missed. Both authors provide sage advice. So, in specifying the time period for behavioral items, you should avoid asking respondents to recall daily or routine events (e.g., visits to the grocery store, exercise class) after a long period has passed, but the time period should be long enough to capture events occurring more rarely (e.g., graduations, births).

Guidelines for Items about Demographics (Personal Background)

Researchers frequently examine the relationship between attitudes, behaviors, and knowledge and the demographics of survey respondents. Items that address demographics collect information about respondents' backgrounds (Fink, 1995). Types of demographic information frequently collected include age, gender, income, marital status, ethnicity, and education level.

A first rule about demographics is to only ask about those characteristics that are relevant to the research questions. Respondents will at times balk at providing estimates of their income or their age. Thus, in asking for personal information, you might lose some of your respondents. So, ask yourself whether information about respondents is critical to the research.

When asking about age, ask for the precise date of birth rather than current age. Respondents' recall of current age is sometimes slightly off; so, asking for their birth date avoids this issue.

Income levels should be relevant to the respondents. Incomes of patients at a free clinic will differ from those at a private practice. To find appropriate income ranges, ask those who work with the respondents the

possible income levels. Also, the time period requires consideration; thus, ask about *annual* income or *monthly* pay.

In deciding about labels to use for your demographic group, use the classification preferred by the groups responding to your survey. The *Publication Manual of the American Psychological Association* (APA, 2010) provides guidance on the use of labels. In addition, the list of ethnic groups used by the Census Bureau provides guidance in specifying demographics.

Number of Items

To this point in developing items for a survey scale, we have discussed the need to consider the relevance of the items, the audience, language, item structure, and conventions. You might ask, “How many items do I need to develop?” Fabrigar and Wood (2007) wrote that “attitudes are often multifaceted, involving cognitions, emotions, and behavioral tendencies. A single item is unlikely to capture the full scope of the attitude in question; using multiple items potentially ameliorates this problem” (p. 537).

The number of items depends on what the researcher needs to know and on how many items are needed for accurate measurement (Fink, 1995). In terms of what the researcher needs to know, the survey on transition contained six subscales: Student–Teacher Interactions, Teaching Styles, Grading Practices, Participating in Class, Getting Along with Others, and School Climate. Each subscale had between 6 and 14 items. As we shall learn in Chapter 8, to attain stable scores on a scale, the more items you have, the better. However, the number of items must be balanced with the amount of time the respondents will dedicate to the survey.

In this chapter, we focused on writing survey scale items to which study participants will respond. In Chapter 5, we discuss the choices to be made in selection of the item response scales.

FURTHER READING

Fink, A. (2003). *How to ask survey questions* (2nd ed.). Thousand Oaks, CA: Sage.

Provides guidelines for writing survey items. Includes examples to illustrate the application of guidelines.

Peterson, R. (2000). *Constructing effective questionnaires*. Thousand Oaks, CA: Sage.

Presents an in-depth discussion of item-writing guidelines.

CHAPTER EXERCISES**1. What revisions are needed for the item below?**

My teacher uses many types of instruction activities (lecture, small-group work, computers).

2. In the item below, how should the phrase “special-needs students” be revised?

I need additional materials to teach special-needs students.

3. What is the reading level of the item below? Is the reading level appropriate for the audience?

My supervisor is responsive in addressing problems that affect my job.

4. Review the item-writing guidelines and develop an example that shows (a) the violation of a guideline in the item and (b) the same item following the guideline. Use the NOT THIS/BUT THIS form to create your example.

<input type="radio"/> NOT THIS
<input checked="" type="radio"/> BUT THIS