Measuring Affective Components of Listening

Chapter · August 2017

DOI: 10.1002/9781119102991.ch5

CITATION
1

READS
411

2 authors:

Graham D Bodie
University of Mississippi
121 PUBLICATIONS **1,805** CITATIONS
SEE PROFILE

Susanne M. Jones
University of Minnesota Twin Cities
42 PUBLICATIONS **1,111** CITATIONS
SEE PROFILE

5

Measuring Affective Components of Listening

Graham D. Bodie and Susanne M. Jones

Listening competence is a function of three interrelated components: knowledge of listening, the motivation to listen, and appropriate and/or effective performance when listening (Coakley, Halone, & Wolvin, 1996; Halone, Cunconan, Coakley, & Wolvin, 1998; Halone, Wolvin, & Coakley, 1997). The second of these components, *motivation*, is most often defined as an "attitudinal component—the willingness to engage as a communicating listener" (Wolvin & Coakley, 1994, p. 151). Motivation characterizes the affective relationship between speaker and listener (Cronen & Price, 1976), regulating whether people approach or avoid one another (Elliot, Eder, & Harmon-Jones, 2013; Gollwitzer, Fujita, & Oettingen, 2004; Weiner, 1992).

The need for listeners to develop a "positive listening attitude" is pervasive in current definitions of effective listening (see Chapter 1), and the importance of such an attitude is stressed in listening instruction across K–12 (Cooper, 1998) and higher education (Wolvin & Coakley, 2000) alike. Most textbooks stress the importance of "knowing why you are listening" and being aware of listening-related goals and priorities (Brownell, 2013; Worthington & Fitch-Hauser, 2012). In particular, students are taught to take responsibility as a listener, to "attend to others with an open mind," and to recognize "that listening is an active not passive activity" (Wolvin, 2009, p. 137). Each piece of advice taps some facet of the affective dimension of listening with the general assumption that sheer knowledge is not enough to listen well: The genuine desire to listen effectively is equally important.

In addition to being a popular component of teaching listening, affective components of listening also are popular in the academic literature. Keaton and Bodie (2013) reported that 80 out of 110 studies published in the *International Journal of Listening* (*IJL*) between 1987 and 2011 (nearly 75%) examined one or more facets of motivations to or tendencies toward listening. Mirroring work in the social sciences more generally, most of these studies ($n = 67$, 61%) asked participants to report on their own attitudes, motivations, or perceived tendencies; the remaining ($n = 13$, 39%) asked participants to report on another person (e.g., a friend, coworker, or spouse). As a result, much of what we know about listening is limited to what people report about their own listening (retrospective self-report) and how this self-knowledge aligns with (or diverges from) what other people report (retrospective other-report).

This chapter begins by discussing the strengths and limitations of self-reporting methods. After reviewing how self-reporting methods have been used (and abused) in the listening literature, we outline a set of recommendations for proper use.

The Sourcebook of Listening Research: Methodology and Measures, First Edition.
Edited by Debra L. Worthington and Graham D. Bodie.

A subsequent section discusses several popular measures of affective components to illustrate how research employing them can best be interpreted. The final section, "New Horizons," details how listening scholars can measure affective elements of listening in more nuanced and potentially powerful ways.

Self-Report Measurement

Listening researchers have two primary sources of information at their disposal from which they can make claims about listening: observation and reporting. Methods appropriate for capturing listening behaviors are covered in Chapters 3 and 6. This chapter deals with affective data that capture internal states reported by research participants, which is referred to as *retrospective self-report* or *self-reporting*. Although self-reports also are used when asking people to judge others (e.g., "How well do you think your friend listens?"), we focus here on utilizing self-reports to capture one's own perceptions.

Participants can report on any number of listening variables, including "recollected behavior, experience, events, and affect, as well as global assessments of affective/psychological states and typical behavior based on accumulations of previous experience and knowledge" (Metts, Sprecher, & Cupach, 1991, p. 162). Each type of reporting method has been used to make claims about listening, although a significant portion of this work has focused on discovering trait-like dispositions thought to affect how people behave as listeners. Traits or dispositions are "characteristics of people that are relatively stable across time and situations" (Hoyle & Leary, 2009, p. 12); situational fluctuations or more transitory reactions are captured with state measures. In many cases, trait and state measures differ only on the basis of the specified time frame. Stable (trait) tendencies are captured by asking participants to evaluate "in general" (e.g., "In general, how well do you listen?"). Situational (state) listening measures refer participants to specific situations and times (e.g., "How well did you listen to your partner in the last conversation?").

The earliest self-report measure of a listening trait was the Receiver Apprehension Test (RAT). Wheeless (1975) originally defined Receiver Apprehension (RA) as "fear of misinterpreting, inadequately processing, and/or not being able to adjust psychologically to messages sent by others" (p. 263). RA has since evolved into a construct called Informational Reception Apprehension (IRA), a three-dimensional construct related to an individual's anxiety regarding: (a) listening, (b) reading, and (c) thinking about abstract concepts (Wheeless, Preiss, & Gayle, 1997). We use the Informational Reception Apprehension Test (IRAT) in this chapter to illustrate the advantages and disadvantages of self-reports of listening (see IRAT, Profile 27).

Advantages and Disadvantages of Using Self-Reports

As seen in Table 5.1, each item of the listening subscale was written to reflect how the respondent typically feels while listening. Only the last item ("I have avoided listening to abstract ideas because I was afraid I could not make sense of what was said") requires the respondent to explicitly think about a past behavior (something he or she did); all other items reference experiences or feelings during the interaction. These items do *not*

Table 5.1 Items comprising the IRAT-Listening Subscale.

While listening, I get nervous when a lot of information is given at once.
I get impatient and anxious when listening to someone discuss theoretical, intellectual issues.
I feel agitated or uneasy when someone tells me there is not necessarily a clear, concrete way to deal with an important problem.
While listening, I feel tense when I have to analyze feelings carefully.
When I hear abstract material, I am afraid I will be unable to remember it very well.
It is frustrating to listen to people discuss practical problems in philosophical and abstract ways.
Many classes are annoying and uncomfortable because the teacher floods you with detailed information in the lectures.
I experience anxiety when listening to complex ideas others tell me.
When I listen to complicated information, I often fear that I will misinterpret it.
I feel relaxed and confident while listening, even when a lot of information is given at once. (R)
Listening to complex ideas is a pleasant, enjoyable experience for me. (R)
When listening, I feel relaxed and confident that I can remember abstract ideas that are being explained. (R)
I have avoided listening to abstract ideas because I was afraid I could not make sense of what was said.

Notes: Scale items reprinted with permission. Items marked with (R) are reverse-coded prior to computing an individual's listening anxiety score. For more information about how to score listening anxiety, see IRAT Profile, 27.
Source: Wheeless *et al.* (1997).

capture observable behavior that can be easily witnessed and reported on by another person (i.e., a trained rater or a participant reporting on listening behavior). These items *do* capture affective components, elements of listening that are likely best reported by the individual listener. Of course, quite a few of the IRAT items capture emotions that may be "visible" in a person's behaviors, such as blushing or shaking, which would then make them observable. Emotional experiences are not, however, isomorphic with emotional expressions (Ekman & Friesen, 1975).

The actual experience of nervousness is an internal state known only to the person who experiences it. As McCroskey (1997) argued, "self-report measures … are most appropriate when they are directed toward matters of affect and/or perceptions in circumstances … [and] least useful when they are directed toward matters of fact … unknown or unknowable by the respondent" (p. 196). We recommend that research questions measuring some form of listening motivation utilize self-report scales. Motivational states are affective in nature, and internal states are best assessed by those who experience these states.

In addition to capturing information that may not be readily available from other methods, there are several other reasons why a listening researcher may want to utilize a self-report measure. First, self-report measures are easy to administer and generally inexpensive, although some scales are not free of charge (e.g., the Doctors' Interpersonal Skills Questionnaire, Profile 15). Researchers who have access to online software such as Qualtrics or SurveyMonkey can populate a survey within minutes of Institutional

Review Board (IRB) approval. Moreover, because most self-report research is exempt from full review by university IRBs, they reduce time-consuming paperwork and meetings in comparison with designs that are more intrusive (e.g., deception studies). Second, when administered online, researchers are able to capture data from captive (e.g., college students and paid survey panelists) and noncaptive (e.g., genuinely interested constituents) samples by posting the URL on bulletin boards, listservs, or mobile data collection labs. When utilizing college students, online administration can be advantageous because students may be more likely to participate if they are not required to report to a laboratory at a predetermined time of day. If your research requires non–college student participants, online administration may be the only logically feasible way to collect data. Third, when properly developed, self-report scales often produce scores with high levels of reliability.

Although advantageous for many reasons, there are limitations to the use of self-report measures. First, self-reporting relies on respondent honesty, an issue that varies as a function of the level of stigma, taboo, or sensitivity of the topic. The desire to create a favorable impression might be more pronounced for some participants or in some contexts compared to others (e.g., when listening assessments are part of performance appraisals). Social desirability of a socially positive and acceptable behavior like listening can cause people to overreport tendencies or motivations (Crowne & Marlowe, 1960; Lawson & Winkelman, 2003). For example, students might be honest about their willingness to listen in a classroom setting when asked by a researcher who is distributing an anonymous and confidential survey outside of that setting. If asked these same questions by their instructor, however, students may feel pressure to answer in a way that gives the impression they are attentive to class material even if the survey guarantees complete anonymity. If social desirability varies significantly as a function of group membership or the situation, then the researcher is faced with a systematic source of error (response bias) that is difficult to manage. Even in research contexts that do not seem prone to social desirability effects, respondents may be over- or underreporting; that is, their perceptions and thus their judgments of their own internal states are influenced by desires (e.g., "I *want* to be more sensitive to what others say") and culturally scripted norms and expectations (e.g., "I *should* be more sensitive to what others say"), rather than their "true" internal tendencies (e.g., "I *am* more sensitive to what others say"). Of course, it is always the case that other biases or moods cloud judgments of one's own internal states. In fact, people cannot introspectively assess themselves in a completely accurate manner; that is, participants may be less self-aware than researchers assume (Nisbett & Wilson, 1977). But the social desirability bias adds just another source of systematic error.

Second, even when social desirability is not a major concern, participants may not have the ability to provide accurate judgments about their internal states. Measures that assess facets of attitudes or motivation that participants have not previously considered are particularly prone to this problem, as are measures designed to assess retrospectively recalled attitudes (e.g., "How nervous were you last week?"). Unfortunately, participants readily provide judgments for attitudes, feelings, and past experiences they have never experienced or pondered, a problem further compounded by online surveys that often force participants to respond to all scale items. Most surveys do not include a scaling option for "I have never considered that" or "I do not know." Even if these options were available, some may be reticent to choose them for fear of looking incompetent or otherwise unable to make a decision.

When asking people to report on behaviors, they may not be aware of how they typically act or how they did act, "especially if they are asked about experiences from the distant past or things that are not very salient" (Metts *et al.*, 1991, p. 168). The ability to remember specific details of conversations or events pales in comparison with our tendency to remember the gist of situations or people (Stafford, Burggraf, & Sharkey, 1987; Stafford & Daly, 1984; see Profile 38, Memory for Conversation, as well as Chapter 4). Indeed, researchers tend to overestimate participants' abilities to report accurately on enacted behavior (e.g., the amount of eye contact, or the number of open-ended questions asked).

In our view, it is valid to use self-report scales that capture some perspectives or subjective recall of experiences ("impressionistic") (Metts *et al.*, 1991). When, however, participants are asked to retrospectively recall behaviors, chances are that their recollections will be biased by, for instance, current mood or unique experiences (Miell, 1987). Work showing that reports of behavior are not significantly correlated with actual enactments of listening (Bodie, Jones, Vickery, Hatcher, & Cannava, 2014) suggests that when respondents are asked to report on their own listening competence, scores should be interpreted as (a) *motivations* for or *attitudes* toward taking an active view of listening or (b) an individual's *beliefs* about his or her competence. A researcher using self-report scales of listening should refrain from labeling these kinds of data as reflecting actual behavior (see Chapter 6 and below for further discussion of this issue).

Third, even when respondents are able to report behavior accurately (e.g., they truly can remember, recall, and/or introspect appropriately), they might be unable to understand or interpret particular items. The ability to comprehend items is more likely to affect the measurement of abstract concepts, such as "feeling insecure while listening" as opposed to "my heart beat fast when I had to listen to that lecture" (see Table 5.1). Because it is abstract and uniquely tied to personal experience and personality, concepts like "insecurity" conjure up multiple meanings. When constructing survey items, be cognizant of writing clear directions and using concrete, familiar, and unambiguous language (see Chapter 2).

It is nearly impossible to ensure that all respondents interpret all questions in the exact same way, which is one reason to develop scales cumulatively and continue to assess their psychometric profile. In addition, most scales that assess affective listening components use ordinal-level scales (e.g., from *strongly agree* to *strongly disagree*). Just as respondents interpret items differently, they also interpret and use scales differently; what Person A rates as 4 (*agree*) might be rated as 5 (*strongly agree*) by another person, not because they have different opinions but because they interpret the meanings of scale points differently. Differences in scale point interpretations produce different scores that reflect a source of systematic error and call for questioning the validity potential of the scale. Some work suggests that scale use is a function of one or more personality traits (e.g., Austin, Deary, Gibson, McGregor, & Dent, 1998).

Finally, with the proliferation of online survey software, researchers are faced with choices regarding ease of data collection and a need to balance convenience with control over the survey environment. Although online technologies allow research teams to collect larger, more diverse samples in an efficient manner, these technologies also introduce concerns about the validity and representativeness of data. For instance, researchers have to assume participants are paying adequate attention when completing an online survey. One way to test whether participants are paying adequate attention is to introduce questions that are randomly distributed throughout core items of a specific

scale. Including an item such as "Please click the radio button under the number 4 for this item" in the middle of a survey allows you to remove respondents who clicked any other button.

Conclusion

Numerous self-report measures of listening are profiled in this book, and there is good reason to include one or more of them in your research. Self-reporting is viable whenever you would like to assess one or more affective components of listening, such as motivation, attitudes toward listening, and beliefs about listening. When there is a need to design an instrument to measure a currently unavailable component of listening, we direct you to the general guidelines presented in Chapter 2. We will return to an overview of specific measures toward the end of this chapter. For now, we turn our attention to common uses and abuses of self-report methods in listening research.

The Uses and Abuses of Self-Report Methods in Listening Research

The purpose of this section is to provide an overview of how self-report methods have been used in listening research. We first detail some general tendencies of self-report listening research. A second section then offers a few suggestions for future improvement.

Some General Patterns in Self-Report Listening Research

Our first observation about self-report listening research is that it tends to be cross-sectional in nature; that is, participants are asked to answer several self-report items during a single data collection session. Although cross-sectional, self-report research is useful for many purposes. We, however, must recognize several methodological issues. First, the only conclusions you can draw from cross-sectional research are correlational, rather than causal, in nature. If you want to test cause–effect relationships, you need to show that (a) the cause temporally precedes the effect, (b) both cause and effect are related with one another in meaningful ways, and (c) there are no other variables that could have plausibly caused the effect (Shadish, Cook, & Campbell, 2002). Causal relationships are most difficult to establish in the social sciences, including communication sciences, because there usually are many causes that influence effects. Work on listening anxiety has found that this trait correlates with reports of verbal aggression (Schrodt & Wheeless, 2001) and a variety of other trait-like personality and communication variables (e.g., Ledbetter & Schrodt, 2008; Schrodt, Wheeless, & Ptacek, 2000). What we are left to speculate about, however, is whether listening anxiety is *caused* by these traits, whether these traits are caused by anxious dispositions, or whether some third variable explains the association between anxiety scores and other trait-like variables. Discovering what causes an individual to experience listening anxiety would have great theoretical and practical payoff, yet a definitive answer is impossible by relying on cross-sectional data. Questions of causality are best answered with experimental research or longitudinal studies that track a set of individuals over time.

Our second observation is that self-report scales tend to generate quite a bit of error. Measurement error can be divided into two types: random error and systematic error. *Random error* contains any factor that randomly affects the measurement of a variable across a sample. For instance, some participants in a study on listening anxiety might experience brief lapses of attention or fatigue that interferes with responding. Fatigue artificially inflates or deflates participant scores but does not affect all participants in the same way (i.e., it is not systematic). To put it in statistical terms, random error affects variability around the average of a sample, but it does not influence the average itself.

Systematic error, on the other hand, *does* affect the sample average; as a result, differences among groups can be a hidden function of systematic error rather than of true group differences. *Systematic error* consists of any factor that systematically affects the measurement of a variable across a sample. For instance, a researcher may administer a survey when classes end and when there is considerable noise, whereas other administrations might be done when there is no external noise that interferes with survey participants. Because entire sets (i.e., an entire class) of participants are affected by noise in a similar manner, in this example external noise is a source of systematic error.

Whereas random error, by nature, is impossible to predict and control, there are several strategies one can take to limit systematic error. Consistency in testing environment and instructions is crucial, especially as the number of people collecting data increases; proper training is a must. If there are known influences on a key variable, one way to limit systematic error is to either control those influences (e.g., collect data at multiple times of day) or measure those influences (e.g., track the time the survey is taken); statistical procedures can then be used to test for any systematic error from those influences. One potentially dangerous and yet not highly recognized source of systematic error is known as *common method variance* (CMV). CMV occurs when shared variance among scores is a function of the method used to collect data rather than the constructs the measures represent (Podsakoff & Organ, 1986; Richardson, Simmering, & Sturman, 2009). In the work on listening anxiety, for instance, because respondents self-report both their listening anxiety and their aggressive tendencies, it is possible that reported correlations between these constructs are a function of a common reporting format for these variables and not a function of an actual relationship between them (see Kotowski, Levine, Baker, & Bolt, 2009). Ways to handle CMV are covered in the next section.

A third issue with self-report listening research is that most studies do not present any evidence for construct validity. *Construct validity* is the extent to which a scale measures the theoretical construct of interest (e.g., listening anxiety). Typical in the listening literature are studies that report evidence for convergent and discriminant validity by showing correlations between a newly constructed scale and measures of theoretically relevant constructs, or *nomological network validity* (Cronbach & Meehl, 1959). Nomological networks are representations of connections between concepts, their observable manifestations, and how these concepts are related. What studies often fail to do, however, is provide adequate empirical evidence that justifies the creation of a new construct and its observable manifestation—the actual scale. Questions that remain unanswered include whether we need the new scale (e.g., whether it replicates an existing construct) and whether the items that comprise that new scale actually "represent" the construct of interest. Perhaps most emblematic of this problem are studies that create scales to measure self-reported listening competence. As reported by Fontana, Cohen, and Wolvin (2015), there are dozens of scales that allegedly measure

perceptions of listening competence. Some of these scales have some items in common, but most items are usually not isomorphic. Thus, the researcher who constructed the scale may lay claim to a supposedly "new" kind of listening competence construct, even though there are scales that already measure exactly that, albeit with different items. To date, perceived listening competence scales have not been administered simultaneously to a sample of participants to determine scale overlap. We are left to wonder whether these scales actually do measure different facets of listening competence or whether they are simply iterations of the same construct. The literature can thus become littered with conceptually indistinct and hence unnecessary constructs, a problem known as *construct proliferation*: "the accumulation of ostensibly different but potentially identical constructs representing [listening] phenomena" (Shaffer, DeGeest, & Li, 2015, p. 1). Ways of ascertaining evidence for discriminant validity of a particular scale are discussed in the next section.

Of course, discriminant validity is only one part of the overall psychometric portfolio of a scale. Part of the validity portfolio of any scale is evidence that the scale factors as expected across a range of populations. In other words, if a listening scale contains various subscales, then evidence is needed that items making up each subscale "hang together" in expected ways. A prime example of this problem comes from the Listening Styles Profile (LSP-16; see Profile 36). The LSP-16 has been used widely by researchers, educators, and practitioners, but the factor structure has been assumed as stable in these studies rather than systematically tested and retested. Instead, the scale had only been submitted to principal component analysis, and in those analyses the four listening styles accounted for just over 50% of the variance among the 16 items (Bodie & Worthington, 2010). The remainder of the unexplained variance is explained either by another set of factors or by measurement error. When submitted to confirmatory factor analysis (CFA) using data from more than 700 participants, Bodie and Worthington found that the scale was not psychometrically sound. A large culprit for this poor model fit was random error, suggesting that individual items were poor indicators of their supposed constructs (i.e., the listening styles). New items needed to be written, and a case for construct validity needed to be made anew (see Profile 36). The assumption of much listening research is that if a scale has been published, it is "valid"—a dangerous assumption indeed. Many "established" scales lack validity evidence, but even in the face of no evidence (or counterevidence), people still use them (Levine, 2005). Scales need to stand the test of time and be submitted to rigorous tests of validity. Notably, validity is an ongoing process. Scales are not valid or invalid; instead, scales can have larger or smaller validity portfolios.

A final observation about self-reported listening research is that it is heavily biased toward assessing self-reports of behaviors or behavioral tendencies rather than self-reports of internal tendencies, an issue we discussed in this chapter. Although there are examples of self-report listening scales directed toward internal cognitive states (e.g., IRAT; see Table 5.1), beliefs about listening (e.g., the Listening Concepts Inventory), motivations (e.g., Willingness-to-Listen), and situationally derived goals (e.g., LSP-R), there are many more examples of scales attempting to gauge behaviors. As described in Chapter 6, *behavior* is defined as something a listener *does*—specific actions a listener enacts while interacting with another person.

Examples of scales that attempt to measure behavior through self-report (all are profiled in Section Three) include the Organizational Listening Survey (OLS), the Self-Reported Listening Competence (SRLC) scale, the Active-Empathic Listening Scale

(AELS), and the Academic Listening Self-rating Questionnaire (ALSQ). Although each scale attempts to measure slightly different facets of listening competence, the scales are similar insofar as they assume people are able to accurately report on what they typically do while listening. Work by Bodie *et al.* (2014) and others (Cooper & Husband, 1993; Ford, Wolvin, & Chung, 2000; Lawson & Winkelman, 2003) questions this assumption; participants tend to overestimate their own listening abilities. In general, the reader is urged not to rely on self-report methods when interested in studying listening behavior.

How to Improve Self-Report Listening Research

In this chapter, we referenced and briefly outlined four abuses of self-reported listening research, namely, that it tends to: (a) Be cross-sectional, (b) contain measurement error, (c) rely on scales that lack validity evidence, and (d) measure behavior rather than internal states. In this section, we discuss some solutions to these abuses.

Make Cross-Sectional Work Part of a Larger Program of Research

A cross-sectional study involves measuring a set of variables at one point in time only. A typical cross-sectional study involves creating a questionnaire by combining multiple scales and demographic information such as age, biological sex, and occupation and distributing that questionnaire online or in person to a sample of individuals drawn from a selected population of interest. Such a design allows researchers to compare many different variables at the same time at a relatively low cost. But because a cross-sectional survey captures only a single moment in time, we are left only with correlational data and must speculate about which variables are antecedent and which ones are consequent. For instance, if you distributed a questionnaire that included the IRAT (see Table 5.1) and a measure of typical communication patterns in families, you might find that individuals scoring higher on listening anxiety also report coming from families that stress a climate of homogeneity of attitudes, values, and beliefs (i.e., conformity orientation; see Ledbetter & Schrodt, 2008). Although it might be tempting to conclude that the ways that families communicate *cause* children to develop a tendency to experience listening anxiety, such a conclusion is beyond the available data.

For questions that require causal answers, longitudinal designs can be useful. Like cross-sectional studies, longitudinal studies involve asking participants to report on their dispositions, typical activities, preferences, and/or perceptions. In a longitudinal study, however, researchers repeatedly measure the same participants over a period of time, sometimes lasting several months or even years. This kind of longitudinal design is called a *cohort* or *panel study*. Another type of longitudinal study covered in depth in Chapter 3 is *ethnographic fieldwork*. Neither panel studies nor ethnographic work allows the researcher to establish causality. Regardless, longitudinal designs allow researchers to explore how key variables change over time and to discover evidence that suggests one causal order might be more plausible than another (e.g., listening anxiety and family communication patterns).

A study design that can more clearly enable causal claims is the experimental study, more specifically, the *randomized controlled trial* (RCT), which is the "gold standard" of evidence-based empirical research. The RCT requires that samples be randomly drawn and that participants be randomly assigned to conditions. Whereas cross-sectional and longitudinal designs allow researchers to investigate issues without directly influencing

or manipulating the environment, experiments involve manipulating the environment and then observing and measuring reactions to those manipulations. The variable that is manipulated is the *independent variable*, and the variable that is thought to change as a function of the manipulation is the *dependent variable*. If researchers are concerned with causal relationships (as we often are in listening), the appropriate design is the experiment, offering full control over the independent variables under question. To study listening anxiety and family communication patterns, for instance, we might train some families to exhibit behaviors more indicative of a conversation orientation, one in which all family members are encouraged to participate in unrestrained exploration of all kinds of topics. Then we would document the degree to which listening anxiety changed as a function of this intervention. If anxiety levels vary as a function of the intervention relative to control groups that did not receive the intervention, we have more definitive evidence of a causal relationship between these two variables. But we still must rule out other plausible variables that could confound our results, and we must be cautious because this particular example did not allow for random assignment, which might, in fact, introduce error (a concern we have addressed in this chapter).

In general, the best method for your own research depends on the questions you are asking. Moreover, your choice of research method is at least partially a function of the research that was previously done on your research question. You must assume researchers before posed precisely the same question and tested it. All scientific research is cumulative, and we encourage you to thoroughly research your question to determine (a) the existing knowledge base and (b) existing scales. Depending on your specific area of listening research, you might be able to conduct a cross-sectional study. Returning to the Ledbetter and Schrodt (2008) report on family communication patterns and listening anxiety, the study was the first to establish whether there are links or associations between these constructs. A cross-sectional study design was thus appropriate because the purpose of their study was descriptive: to discover the prevalence of one or more variables in a particular population or subgroup of that population. When conducting work in an area that has quite a bit of formative research already, however, longitudinal or experimental designs are more appropriate, yet also more costly.

Create New Constructs (and Measures) Only When Necessary

As interest in listening has increased, so too have the number of constructs used to describe its operation. At first blush, this does not seem overly problematic. Indeed, to understand something as abstract and complex as listening, one would assume the need for myriad constructs and operative mechanisms. As the Greek poet Hesiod famously wrote, however, moderation is best in all things. According to Occam's razor, which sets the standard for good scientific theory, when possible, the most parsimonious explanation is preferred. For listening scholars, that means generating the fewest number of constructs possible to explain how and in what ways individuals behave as listeners.

Similar to our discussion in Chapter 1 regarding the proliferation of listening definitions, there also is a tendency to add adjective descriptors to listening in hopes to set these listening constructs apart from already existing constructs. Terms like *active listening* (Rogers, 1955; Weger, Castle, & Emmett, 2010), *therapeutic listening* (Wolvin & Coakley, 1993), *supportive listening* (Bodie, Vickery, & Gearhart, 2013; Jones, 2011), and *empathic listening* (Myers, 2000) all describe a mindful approach to listening, one that truly seeks to understand what others are really saying. Adding to this proliferation are closely related constructs, such as *responsiveness*, *understanding*, *attentiveness*,

conversational sensitivity, *interaction involvement*, and *affectionate communication*, that share a great deal of conceptual space with listening and may not be empirically distinct from it. Then, there are broader terms such as *empathy*, *cognitive complexity*, and *listening anxiety* that likely contribute to an individual's ability to listen in a competent manner, as well as terms such as *listening goals* and *listening concepts* that describe cognitive schemata related to listening.

The basic point is that authors of supposedly new constructs should, at minimum, advance a strong theoretical argument for conceptual distinctions between their "new" construct and related constructs. Advancing an argument forces scholars to examine existing knowledge bases and may encourage them to abandon their constructs, which might ultimately strengthen a research field. Pragmatically, most journal editors and reviewers require authors to advance an argument for conceptual distinctiveness. As explained by Harter and Schmidt (2008), however, the "implicit assumption … [that if] they and other researchers can make a logical or conceptual distinction between constructs or measures, then this distinction will exist in the minds of … respondents to surveys … may not hold true" (p. 36). In other words, constructs that are conceptually distinct may not be empirically discrete when data are collected from surveys created to operationalize these constructs.

Measures that are similar lack empirical distinctiveness and are correlated at or near 1.0 after correcting for measurement error. Evidence from other literatures (e.g., organizational science) suggests that the empirical distinctiveness between related constructs is rarely investigated (Shaffer *et al.*, 2015). Broadly speaking, we are in the realm of *discriminant validity*, or the extent to which measures of theoretically distinct constructs are unrelated empirically to one another. Gathering evidence of discriminant validity is part of generating strong evidence for construct validity.

Generate Strong Evidence for Construct Validity

As mentioned, current listening measures often exhibit a good deal of measurement error and lack sufficient evidence for construct validity. Measurement error is unavoidable, of course, because the constructs we wish to study are abstract idealizations of a concrete empirical reality. And, of course, no measure ever fully captures the construct that it operationally defines—a frequent and misguided assumption known as *definitional operationalism* (Campbell, 1969). But sound research design and strict adherence to data collection procedures can eliminate a lot of error. Procedures outlined in Chapter 2 can assist in the development of measures that include lower levels of error.

A very basic way of modeling measurement error is to look at Cronbach's alpha as an estimate of the internal consistency of a set of scores. Chapter 2 presented the formula for alpha, and this statistic can be used to explore how much measurement error has attenuated the relation between two constructs. Say, for instance, we have collected responses from both the Self-Perceived Listening Competence Scale (SPLCS; see Profile 57) and the listening anxiety (LA) items from Table 5.1. Furthermore, suppose we find a correlation of .60. Suppose, in addition, that the reliability estimates for the SPLCS and LA are .76 and .68, respectively. The formula used to correct the observed correlation for measurement error is:

$$r_{corrected} = \frac{r_{xy}}{\sqrt{(r_{xx})(r_{yy})}}$$

where r_{xy} is the Pearson product moment correlation coefficient, estimating the SPLCS–LA relationship; r_{xx} is the internal consistency estimate of SPLCS; and r_{yy} is the internal consistency estimate of LA. Plugging in our values, we get a value of $r_{corrected} = .83$. This value is not 1.0, but it is rather high, suggesting quite a bit of overlap between the SPLCS and LA. This correlation does not suggest that LA and listening competence are the same construct, but it does suggest that the scales used to measure these constructs are not as distinct as the developers originally thought.

Measurement error can also be accounted for and modeled through CFA, which also allows for testing a central facet of construct validity, the degree to which a scale factors appropriately across a range of populations. Unlike Pearson's coefficient, the bivariate relation generated by CFA will be corrected for measurement error, making the extra step of correcting for attenuation unnecessary. In addition, the CFA output will include factor loading information, as well as information regarding the extent to which the data conform to the theoretical measurement model—that is, the model presented by the test developers (Hoyle, 2000; Levine, Hullett, Turner, & Lapinski, 2006; Raju, Laffitte, & Byrne, 2002; Thompson, 2004).

Building a case for construct validity also requires evidence that the scale is related with theoretically similar measures (convergent validity) and lacks associations with theoretically dissimilar measures (discriminant validity). The former evidence is more common in the listening literature and generally consists of authors developing a nomological network for the construct under question and then including measures of the associated constructs along with the new measure. For instance, if we were developing a scale of LA, we might also ask participants to complete scales or tasks that tap listening comprehension and listening competence, reasoning that people with higher LA should have lower comprehension scores and lower levels of listening competence. If correlational data indeed were to show positive associations between LA, competence, and comprehension, our LA scale would possess some convergent validity.

The case for discriminant validity is not as readily made in the listening literature. To establish a case for discriminant validity, a first step is to show that our newly created measure does not duplicate existing measures. As seen here, we can do this by correcting the observed correlation for measurement error. We also can use CFA to model items in various configurations to generate a best fitting model that best captures the factor structure of our scale items. In our SPLCS–LA example, we would first specify that the SPLCS items all loaded on that factor and the LA items all loaded on a separate factor. We would compare this *unconstrained model*, the model that allows these constructs to freely covary, to a *constrained model* that specifies that the items are better represented by a single factor (a unidimensional construct). If the unconstrained model produced better fit than the constrained model, we would argue that there is evidence of discriminant validity (for details on how to conduct such an analysis, see Byrne, 2010).

Another method for gathering discriminant validity evidence is known as the *multitrait-multimethod matrix* (MTMM). The MTMM approach was introduced by Campbell and Fiske (1959) as an empirical solution to construct validity. In this approach, researchers employ measures of at least two traits (constructs) using at least two methods. For instance, in their MTMM study, Bodie *et al.* (2014) were interested in two constructs, active-empathic listening (AEL) and nonverbal immediacy (NVI). They measured these two constructs in three ways: self-reports, partner-reports, and behavioral observations. Participants reported on their tendencies to enact both AEL and

Table 5.2 Multitrait-Multimethod Correlation Matrix

	Active-empathic listening				Nonverbal immediacy			
	Self-report	Other-report	Partner-report	Behavior	Self-report	Other-report	Partner-report	Behavior
AEL self-report	**.87**	.23	–.14	–.07	*.50*	.08	.02	.13
AEL other-report	.20	**.89**	.06	.00	.24	*.48*	.01	.35
AEL partner-report	–.12	.05	**.90**	.35	–.01	.15	*.75*	.09
AEL behavior	–.06	.00	.30	**.80**	.00	.17	.28	*.42*
NVI self-report	*.43*	.21	–.01	.00	**.84**	.37	.11	.19
NVI other-report	.07	*.42*	.13	.14	.32	**.87**	.19	.23
NVI partner-report	.02	.01	*.63*	.22	.09	.16	**.78**	.16
NVI behavior	.11	.30	.08	*.34*	.16	.19	.13	**.81**

Note: AEL = active-empathic listening; NVI = nonverbal immediacy. Correlations above the MM diagonal are corrected for measurement error, whereas those below the MM diagonal are not.
Color coding key:

Monotrait-monomethod correlation	Monotrait-heteromethod correlation	*Heterotrait-monomethod correlation*	Heterotrait-heteromethod correlation

Source: From Bodie *et al.* (2014). Table reproduced with permission.

NVI (self-report) and were then paired with another individual who disclosed a stressful event. The two people talked for 5 minutes, after which the disclosers reported on the listeners' AEL and NVI (partner-report). Because the conversations were video-recorded, the conversations were additionally coded for actual AEL and NVI (behavioral observations). In the end, each participant had three scores for AEL and three scores for NVI, which generated a matrix of correlation coefficients (see Table 5.2).

The first type of correlation generated by an MTMM analysis is the *monotrait-monomethod* (MM) correlation. This correlation is synonymous with the reliability coefficient of a scale. When these values are high, the measures are said to demonstrate a high degree of internal consistency. To generate MM correlations, our research team calculated Cronbach's alpha values for each scale (see Chapter 2); we also corrected subsequent correlations using these estimates (as discussed here). The second type of correlation, the *monotrait-heteromethod* (MH) correlation, represents the association between different measurement methodologies used to measure the same construct. In the AEL–NVI study, the two constructs were measured as self-report, partner-report, and observed behavior. The set of MH correlations for AEL showed the extent to which self-reported, partner-reported, and observed AEL correlated with each other (and similarly for NVI). When MH correlations are sufficiently large, researchers are provided direct evidence of convergent validity. Often referred to as *validity coefficients*, these values should also be sufficiently larger than the heterotrait correlations to

demonstrate evidence for divergent or discriminant validity. In the AEL–NVI study, none of the MH correlations were higher than .30, and some were negative, suggesting little evidence for convergent validity. For instance, the correlation between self-reported AEL and partner-reported AEL was not statistically different from zero, suggesting that how the participant saw him or herself was not related to how the partner perceived his or her listening.

The last two correlations produced by an MTMM analysis are both heterotrait correlations, in our case correlations between measures of AEL and NVI. *Heterotrait-heteromethod* (HH) correlations are associations between different measurement methodologies used to measure different constructs, such as correlations between self-reported AEL and partner-reported NVI or observed AEL and self-reported NVI. Second, *heterotrait-monomethod* (HM) correlations are associations between different constructs measured by a common methodology (e.g., self-reported AEL and NVI, or partner-reported AEL and NVI). In the AEL–NVI study, all HM correlations were quite high, suggesting that the method used to measure these constructs explains a good deal of the shared variance among the constructs.

In general (and back to why we started talking about MTMM in the first place), comparisons between the two heterotrait correlations (HH and HM) are used to assess CMV. When the HM correlations are larger than the HH correlations, measurement bias is a concern. Basically, if most of the variance in the dataset is attributable to method, then researchers should be wary of using cross-sectional, self-report data to make definitive conclusions about relations among facets of listening and other trait-like variables. As we showed with the AEL–NVI study, CMV is likely a concern of self-reported listening research using the AELS and NVI measures that should not be ignored. The extent to which other listening constructs might suffer from CMV needs to be examined.

Acknowledge What Self-report Scales can Measure

The discussion of CMV moves us to our final recommendation regarding self-reported listening research: Do not use these measures as proxies for behavior. Appropriately, this discussion also brings us full circle to a key position of this chapter: Self-report measures of listening are most appropriately described as tapping affective components of listening, such as motivations and attitudes toward listening. A listening behavior is something listeners do, not something they think they do, something they remember doing, or even something they think they tend to do in general. Measuring behaviors requires that people observe listeners in action.

In the AEL–NVI study, the data revealed a large degree of CMV. What this means is that individual reports of tendencies to listen in active-empathic ways are moderately correlated to their reports of tendencies to behave in more or less nonverbally immediate ways; the same is true for actual AEL and NVI behavior (as rated by a trained judge), and perceptions of AEL and NVI by a conversational partner. More important, the correlations between methods were much higher than correlations between self- and partner-reported and rated AEL and NVI behaviors. In a larger sense, these results suggest that studies investigating validity issues with single-method measures of listening (e.g., using all self-report measures to demonstrate convergent validity) may be reporting spurious associations rather than real relations among latent constructs (Podsakoff & Organ, 1986). As a result, previously reported associations may need to be adjusted downward to account for this bias.

No matter its degree, the consequences of CMV are considerable. In egregious cases, estimates may be within interpretable bounds yet be entirely a function of shared

methods across constructs that inflate Type I error (i.e., the "glop" problem; Bank, Dishion, Skinner, & Patterson, 1990). The question then becomes how one can minimize (rather than completely eradicate) this systematic error variance. The reader is referred to Richardson *et al.* (2009) to determine appropriate solutions for CMV. At the very least, the degree to which CMV changes the conclusions regarding past work should be addressed (e.g., the degree to which relations among listening and relationship outcomes change as a function of decoder perspective used to operationalize listening).

Self-Reports of Listening: What Are They Good For?

Unlike Frankie Goes to Hollywood's answer to "War and what is it good for?" ("Absolutely nothing!"), we think that self-reports of listening are good for many things. Indeed, we started this chapter by discussing some of the advantages of self-reporting of listening, and we retain that attitude—that, when used appropriately, we can learn much from self-reports of listening. In particular, researchers are encouraged to use self-report measures to investigate internal states, beliefs about listening, motivations to listen in particular ways, and situationally influenced listening goals. We touch on each of these in this section, using scales profiled in Section Three as examples.

Measuring Internal States

Recall that listening scholars have two sources of information available when measuring listening facets: things participants report and things that can be observed. Each source of information can provide data about internal states. Observed physiological markers can be used to measure stress or anxiety, for instance, but for the most part, listening scholars have relied on self-report methods to ascertain how listeners think and feel about listening. Moreover, as we noted in this chapter, for some constructs, self-report may be the only feasible way to collect data.

The IRAT (see Profile 24) measures one internal state related to listening, the degree to which a listener experiences anxiety (see Table 5.1). Examples of other profiled measures that capture internal listener states include the Affectionate Communication Index (ACI), Attributional Complexity (AC), the Interpersonal Reactivity Index (IRI), and the Rational-Experiential Inventory (REI). Each of these measures can be used to tap trait-level characteristics. Most measures also can be modified to tap state-level characteristics, such as the degree of listening anxiety experienced before, during, or after a particular listening event.

Measuring Beliefs About Listening

What people believe about listening can have powerful effects on how they enact (or fail to enact) behaviors in the service of attending to others. Our beliefs about listening also likely influence how we judge others as they listen to us. When asked, participants readily list a consistent set of behaviors associated with good listening (Bodie, St. Cyr, Pence, Rold, & Honeycutt, 2012), most of which are represented on scales that tap self-perceived listening competence.

Two examples of listening competence scales profiled in Section Three are the Self-Perceived Listening Competence Scale (SPLCS) and the Organizational Listening Survey (OLS). These scales are intended as self-reports of how well one listens in

general; or, if instructions are modified, in a context of interest (e.g., in the classroom or at work). As defined in this book, however, behaviors are things people do, not things people think they do; they are concrete actions that are displayed in the moment. As memory research teaches us, we do not retain much specific information after an interaction, especially information about the exact proportion of time we spent engaging in eye contact or the exact number of open-ended questions we asked. Others have used these competency scales to measure perceptions of interlocutors regarding a listener; so, for example, your coworkers could fill out one of these scales with respect to how they think you listen. This strategy, known as *other-reporting*, is valid insofar as you are interested in what other people think about you. But other-report measures do not tell you what another person did, only what the perceiver thinks that individual did. Of course, perception is one form of reality, so there can be power in studying what other people think about certain listeners. One measure designed to assess what others think about a listener's degree of responsiveness is the Perceived Partner Responsiveness Scale (PPRS).

Another strategy for exploring what people think about listening is to ask them to define the term, a strategy used to develop the Listening Concepts Inventory (LCI). The first step in developing the LCI involved exploring the lay and scholarly literature related to listening and gathering several dozen terms considered synonymous with or closely related to listening (e.g., *understanding*, *attention*, and *learning*). Then, college students in the United States and Germany rated the degree to which each of these terms is identical to or not at all similar to listening (see Profile 32 for exact scale points). Results suggested four broad ways in which people think about listening: as organizing information, as relationship building, as learning and integrating information, and as a critical activity. The work on the LCI is similar in many respects to work on implicit theories of relationships (Knee, 1998), personality (Krzystofiak, Cardy, & Newman, 1988), communication (O'Keefe, 1988; also see the Communication Functions Questionnaire, Profile 11), and other facets of human life that influence how we behave in the presence of others.

Measuring Motivations to Listen

The motivation to listen is an integral part of listening competence—in order to behave in effective and appropriate ways, the listener must not only know how to behave but also have the motivation to behave in that way. One set of measures profiled in Section Three was explicitly designed to tap motivations to listen in particular ways. The Willingness to Listen (WTL) scales were designed to directly measure individual motivations to listen to others in various settings; they have not, however, been used extensively in the literature and do not have powerful validity portfolios.

Other scales also tap listening motivation and have slightly more robust validity portfolios, although they are most often positioned (much like measures of listening competence) as measures of dispositions or tendencies to listen in particular ways. Those include the Active-Empathic Listening Scale (AELS), the Attitude Toward Active Listening Scale (ATALS), the Conversational Sensitivity Scale (CSS), the Interaction Involvement Scale (IIS), the Talkaholic Scale (TAS), and several measures of NVI. The MTMM study detailed in this chapter, which explored reports and observations of AEL and NVI, has already clued you in to our opinion regarding the use of these scales as

proxies for behavior. They are, however, useful to the extent that they might tap the motivation to be a particular type of listener. In the case of the AELS, for instance, perhaps this scale taps the degree to which a listener wants to consciously understand another individual from that individual's perspective. Interpreting the scale in this way is supported by high associations between the AELS and measures of empathy (see Profile 2). Likewise, conversational sensitivity might tap the degree to which a listener wants to be sensitive to both content and relational aspects of speech.

Measuring Situationally Derived Listening Goals

A final affective category to which self-reports seem appropriate is the measure of situationally derived listening goals. The first conceptualization of listening-related goals was developed by Watson, Barker, and Weaver (1995), who proposed the construct of *listening style* as the variability in how people attend to and process information. In particular, Watson *et al.* identified four listening orientations—people, action, content, and time—that individuals habitually use, especially in novel situations (Imhof, 2004). Problems encountered in studies utilizing the LSP-16 (Bodie & Worthington, 2010) led Bodie, Worthington, and Gearhart (2013) to revise and frame this typology as representing four distinct "goals that listeners have when engaged in situations that call them to be a particular kind of listener" (p. 17; see LSP, Profile 36). In a similar manner, the typology of listening competencies outlined by Wolvin and Coakley (1993) that directed the development of the SPLCS also can be interpreted as identifying different goals that listeners might seek to accomplish in interaction. Thus, the SPLCS, instead of being framed as a measure of listening beliefs, might best be framed as a measure of the goals available while listening.

New Horizons: Daily Diaries and Experience Sampling Methods

So far, we have covered advantages of self-report measures, as well as proper uses and common abuses of this method. We also detailed some ways to improve self-reported listening research. One key to improvement is to begin shifting our scholarship from an overreliance on cross-sectional self-reports and to incorporate measures with strong validity portfolios into studies that will allow us to explore the ways in which affective components of listening vary over time. One of the best ways to do so is with diary studies.

What Is a Diary Study?

A *diary study* involves asking participants to repeatedly submit self-reports of "events, reflections, moods, pains, or interactions near the time they occur" (Iida, Shrout, Laurenceau, & Bolger, 2012, p. 277). Learning about others from diaries is, of course, nothing new. Historians and literary scholars utilized diary records long before communication studies emerged as a discipline, and the everyday experiences of people have long been an interest of psychologists, anthropologists, linguists, and many other social scientists. Although not classified as a diary method, the ethnographic methods

covered in Chapter 3 have much in common with what we are discussing here—a desire to capture life-as-lived as close to its occurrence as possible and within the frame-of-reference of participants. One main difference between diary studies and methods such as open-ended journaling and participant observation, however, is that diary studies employ standard self-report instruments to maintain a degree of standardization. When participants are asked to report using a standardized form, the method is often referred to as *experience sampling* (Larson & Csikszentmihalyi, 1983).

Of course, standardization does not mean that all diary studies employ a standard set of methods. Indeed, diary studies have employed a range of methods and procedures. The larger point is that the same self-report scales used to capture snapshots of participant internal states, motivations, habits, beliefs, and other affective components of listening can be used in a continuous format that allows a broader (and approximating a motion) picture of the lives of listeners (see Chapter 3). As stated by Shiffman, Stone, and Hufford (2008), "global, retrospective reports … [miss] the dynamics of life as it is lived, day-to-day, hour by hour" (p. 3).

Experience sampling is actually one of the earliest methods used in listening research (Rankin, 1926; see Time Studies, Profile 60). Time-use studies attempt to estimate the amount of time people spend doing various activities, usually breaking down estimates for various 24-hour periods and reporting percentages of time spent on specific tasks. Rankin's (1926) study asked a convenience sample of 21 people to keep a log of their communication activities for one or more days in 15-minute increments from 6:00 a.m. to midnight. Several other studies have been published that sought to replicate and extend these results for specific populations (e.g., college students, scientists, and engineers). Moreover, studies exploring time spent listening highlight the variability in methods that can be employed in diary studies. For instance, studies prior to 1980 primarily utilized one or more forms of *time-sampling* procedures, asking respondents to report at various times of the day what communication activities they were engaged in. Rankin asked respondents to report every 15 minutes, and these logs of time spent were recorded for between 1 and 18 days. Perras and Weitzel (1981) used a similar method with reports every 30 minutes of waking time. Bird (1953) reported having students keep "a running record of minutes spent" in the four modes, whereas Weinrauch and Swanda (1975) asked respondents "to keep a careful record of their time spent in communication" (p. 27). Hinrichs (1964) used a primitive form of *signal-contingent recording*, asking participants to set a wristwatch alarm at five random times during the day and report on their communication up to that point in time. Since the publication of the time study by Barker, Gladney, Edwards, Holley, and Gaines (1980), participants are most often asked "to think back over the last 24 hours and answer the questions based on [this] reflection" (p. 103).

Why Conduct a Diary Study?

There are both methodological and theoretical factors that researchers should consider when conducting a diary study. Although retrospective self-reports provide information regarding reconstructed experience or perceptions, summarizing the past from a respondent's current point of view, daily diaries provide information regarding experience as it is lived. Diary data allow researchers to explore presently felt emotions and close-to-real-time reports of ongoing experiences as well as potentially more accurate retrospective accounts of behavior. Conversely, traditional self-report scales better reflect internal,

motivational, affective, and perceptual processes that play a role in how people behave, but they fall short of capturing behavior per se. It is important to note that "both types of data are relevant to understanding human behavior" (Reis, 2012, p. 5). Listening scholars want to know both "what actually happened … [and] how people experience or understand events in their lives, given time to reflect on them" (Reis, 2012, p. 5). Thus, we are not arguing to throw out self-report measures completely and replace them with diary studies. Instead, researchers should use global retrospective reports alongside more frequent reports of daily experience (as well as experimental and observational methods).

Theoretically, diary methods "make available a different kind of information than traditional methods do, information that provides a novel and increasingly valuable perspective on behavior" (Reis, 2012, p. 4). In particular, the type of information provided by daily reports of experience and behavior is more ecologically valid than retrospectively recalled behavior. For one, the collection of data is closer to the occurrence of the behavior—asking participants to reflect on their listening anxiety immediately after a conversation will produce data that better describe anxiety felt in that conversation compared to asking them to reflect days or weeks after that interaction. As a type of external validity, ecological validity is important for the representativeness of data—the closer to the experience we can get, the more we can generalize to similar experiences. Moreover, daily reports of experiences and behavior are always contextual, compared to the acontextual nature of most global reports or the contrived nature of many experimental studies; how and why people listen are at least partially a function of the situations they find themselves in. Daily diary methods allow researchers to take seriously calls to contextualize listening research.

Daily Sampling of Listening Experiences

Ideally, if researchers want to know about behavior—what a person does when listening—they should observe the listener in action. At the same time, some listening behaviors are not easily observed by researchers. For instance, discovering how dual-earner couples (i.e., both individuals work full-time outside of the home) listen to each other as they talk about the events of their day poses several logistical puzzles, not the least of which is recruitment. Although you might want to observe couples as they engage in this type of talk, how? Do you get them to come to your lab space? If so, when they get there, how do you introduce their task, to talk about the hassles that made up their day? Don't get us wrong: Having couples talk about stressful events in the lab is possible and oftentimes a desirable design choice (see Bodie, Cannava, Vickery, & Jones, 2016; Bodie, Vickery, Cannava, & Jones, 2015; see Couples Helping Exercise, Profile 14). The point is that recruitment for laboratory studies can pose serious logistical inconveniences, costs that have to be weighted in terms of the benefits that experimental methods provide.

Even if you are able to recruit a sample of couples who agree to come to the lab after work, collecting those data is extraordinarily labor-intensive. Not only are you limited to when you can collect these data (i.e., during the time of day the couples usually talk about their days), but also dual-earner couples may be reluctant to spend one or more hours in the lab after a long day of work. Moreover, those with children will require extra compensation or may be particularly reluctant to give you perhaps the only part of their day they get to see their kids awake. As a result, your sample will be limited to those participants willing and able to come to the lab; these people may or may not adequately represent the larger population to which you want to generalize.

A general rule to research by is to match your methods with your questions. In this case, our question involves how dual-earner couples listen during conversations about daily hassles—what Jefferson (1980) called "troubles talk." If there are little empirical data on this topic, we are warranted to ask participants to answer self-report questionnaires about this type of talk, perhaps by distributing a survey to a sample of dual-earner couples that asks them to retrospectively report how they act in and feel about these conversations. We could ask about their general tendencies in these types of conversations, or we could ask them to remember the most recent troubles talk conversation they had and report thoughts and feelings about how they and their partner listened. Alternatively, we could ask respondents to answer survey items several times—perhaps once a week or more. The point of designing a diary study to explore troubles talk is to get as close to the event as possible and to mitigate recall bias and other problems associated with memory for events. If we decided to go the daily diary route, we are then faced with several subsequent decisions.

Choose a Reporting Method

Much like sampling decisions for participants, researchers using diary methods have to choose how to sample events. In our dual-earner example, the event seems rather straightforward. We could define troubles talk for our research participants as "any conversation in which you and your partner talk about one or more troubles or hassles that happened to you over the course of the day." Our instructions could further specify that each couple is to report on each of these conversations that happened face-to-face for at least 5 minutes, a decision that would remove from consideration texts and email that couples might send throughout their day. This method is known as *event-contingent* sampling, reporting for events predefined by the researcher. We could ask participants to fill out our survey as close to the end of each of these conversations as possible, whenever and wherever they happened during the day. This design choice seems preferable to asking participants to report once per day and to recall all such conversations, unless of course we know from past work that such conversations typically happen only once per day (e.g., at the end of the work day as couples are destressing).

Other sampling choices include *time-contingent* sampling, which is sampling at regular, predetermined intervals, and *signal-contingent* sampling that asks participants to record responses when a signal is sent, for example, through an app loaded on participants' mobile devices or through an email or other messaging platform. Time-use studies tend to use a variation of time-contingent sampling, asking participants to report on every 15-minute interval throughout the day. More specific guidelines and typical schedules for time-contingent sampling are found in Larson and Csikszentmihalyi (1983).

According to Reis and Gable (2000), distinctions between event-, time-, and signal-contingent sampling methods "are not merely procedural details; each protocol is tailored to fit particular operational circumstances and theoretical goals, and findings depend to some extent on the choice of method" (p. 198). The reader is encouraged to consult sources already cited in this chapter, as well as additional sources (e.g., Bolger, Davis, & Rafaeli, 2003; Bolger & Laurenceau, 2013; Reis, Gable, & Maniachi, 2014; Wheeler & Reis, 1991) for more information on sampling-related decisions.

How much do Participants Report?

When events are rare, event-contingent reporting methods are likely the best choice. There is no need to ask couples to report continuously on troubles talk, for instance, if that type of

talk occurs only once per day or a few times per week. For more common events, like listening to music, choices have to be made about how often participants should report. With time-contingent sampling, it is rather standard to have participants report once per day or to report, for instance, once in the morning and once at night. For signal-contingent sampling, the researcher has to choose how many times to signal participants without making the study an inconvenience. For some participants, work or other daily events make it impossible for them to respond more than once or twice per day (e.g., when they are not at work).

Depending on the research question, researchers also have to consider the length of the reporting instrument. Clearly, if participants are reporting two or more times per day, you cannot expect them to fill out several hundred scale items. Many researchers who employ diary methods often employ single-item measures of constructs, opting for efficiency over standard concerns about reliability. For instance, a study investigating the experience of listening anxiety might ask, "How apprehensive were you during this listening event?", rather than employ the entire listening subscale of the IRAT; this question seems to get at the general construct under consideration and thus can be argued to be a valid indicator of listening anxiety.

How do Participants Report?

Regardless of the amount of reporting, researchers using diary methods also need to think about *how* data are collected. Some work employs paper-and-pencil diaries, whereas other work employs online or mobile technology. For time- and signal-contingent sampling, software is available (e.g., Qualtrics or Snap Survey) that allows researchers to preprogram emails, having them sent at regularly timed or random intervals, respectively. If you are using online software, you might want to explore how your survey will appear on tablets, e-readers, phones, and standard computers. Most systems now allow you to program surveys to account for differences in survey appearance, but a good rule is to pilot-test your survey before you begin your main data collection. Finally, in our example study of dual-earner couples, an additional consideration is involved, namely, making sure instructions specify that each individual is to report privately so responses are not contaminated by the other's opinions.

Conclusion

Listening is a common and consequential human activity and a fertile area of research. Understanding how and in what ways to attend to and understand others is imperative for building adequate theories of human interaction and behavior, as well as for teaching others to be productive members of society. The ability to conduct research on listening is, however, contingent on the ability to use methods that produce valid data. For most listening research, scholars rely on self-report survey instruments administered at one point in time. When surveys are used to ask participants about the frequency or duration of listening behaviors, the validity of the results are questionable. Instead, self-report methods are best thought of as ways to discover internal states, motivations, beliefs about listening, and situationally derived listening goals. Alternatively, participants can be asked to report on listening-related variables as close to their occurrence as possible, which is an especially useful method when behavioral observations are either logistically difficult or impossible. Such diary methods were once a staple of listening research and should be used more regularly to understand more than just time spent listening.

Most readers can take solace in the fact that there are always problems with methods, no matter the method chosen. Experimental studies, for all the control and internal validity they offer, often lack external validity (and especially ecological validity). Longitudinal studies, for all the rich data they offer with respect to how experiences change over time, are time-consuming and still rely on participant self-reports. Diary data do not ameliorate concerns about participants misrepresenting themselves. One thing is certain: We need much more research on the reliability of various data collection methods and the degree to which these methods are associated with each other. That much work remains to be done should be comforting, especially to graduate students and early-career scholars. The study of affective components of listening is vital to understanding how and why we listen to others in particular contexts and with particular results. We hope this chapter adds to conducting methodologically rigorous and theoretically sophisticated work.

References

Austin, E. J., Deary, I. J., Gibson, G. J., McGregor, M. J., & Dent, J. B. (1998). Individual response spread in self-report scales: Personality correlations and consequences. *Personality and Individual Differences, 24*, 421–438. doi:10.1016/S0191-8869(97)00175-X

Bank, L., Dishion, T. J., Skinner, M., & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with "glop." In G. R. Patterson (Ed.), *Depression and aggression in family interaction* (pp. 247–279). Mahwah, NJ: Erlbaum.

Barker, L., Gladney, K., Edwards, R., Holley, F., & Gaines, C. (1980). An investigation of proportional time spent in various communication activities by college students. *Journal of Applied Communication Research, 8*, 101–109. doi:10.1080/00909888009360275

Bird, D. E. (1953). Teaching listening comprehension. *Journal of Communication, 3*, 127–130.

Bodie, G. D., Cannava, K., Vickery, A. J., & Jones, S. M. (2016). Patterns of nonverbal adaptation in supportive interactions. *Communication Studies, 67*, 3–19. doi:10.1080/10510974.2015.1036168

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L., & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait-multimethod analysis of three perspectives for judging immediacy and listening behaviors. *Communication Monographs, 81*, 495–523. doi:10.1080/03637751.2014.957223

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. M. (2012). Listening competence in initial interactions I: Distinguishing between what listening is and what listeners do. *International Journal of Listening, 26*, 1–28. doi:10.1080/10904018.2012.639645

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2015). The role of "active listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement. *Western Journal of Communication, 79*, 151–173. doi:10.1080/10570314.2014.943429

Bodie, G. D., Vickery, A. J., & Gearhart, C. C. (2013). The nature of supportive listening I: Exploring the relation between supportive listeners and supportive people. *International Journal of Listening, 27*, 39–49. doi:10.1080/10904018.2013.732408

Bodie, G. D., & Worthington, D. L. (2010). Revisiting the Listening Styles Profile (LSP-16): A confirmatory factor analytic approach to scale validation and reliability estimation. *International Journal of Listening, 24*, 69–88. doi:10.1080/10904011003744516

Bodie, G. D., Worthington, D. L., & Gearhart, C. C. (2013). The Listening Styles Profile (LSP-R): A scale revision and evidence for validity. *Communication Quarterly, 61*, 72–90. doi:10.1080/01463373.2012.720343

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616. doi:10.1146/annurev.psych.54.101601.145030

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford.

Brownell, J. (2013). *Listening: Attitudes, principles, and skills* (5th ed.). Upper Saddle Ridge, NJ: Pearson.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Erlbaum.

Campbell, D. T. (1969). Definitional versus multiple operationalism. *Et Al., 2*(1), 14–17.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. doi:10.1037/h0046016

Coakley, C. G., Halone, K. K., & Wolvin, A. D. (1996). Perceptions of listening ability across the life-span: Implications for understanding listening competence. *International Journal of Listening, 10*, 21–48. doi:10.1207/s1932586xijl1001_2

Cooper, L., & Husband, R. L. (1993). Developing a model of organizational listening competency. *Journal of the International Listening Association, 7*, 6–34. doi:10.1080/10904018.1993.10499112

Cooper, P. (1998). *The speaking, listening, and media literacy standards and competency statements for K–12 education*. Washington, DC: National Communication Association.

Cronbach, L., & Meehl, P. E. (1959). Construct validity in psychological tests. *Psychological Bulletin, 52*, 174–203. doi:10.1037/h0040957

Cronen, V. E., & Price, W. K. (1976). Affective relationship between the speaker and listener: An alternative to the approach-avoidance model. *Speech Monographs, 43*, 51–59. doi:10.1080/03637757609375915

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354. doi:10.1037/h0047358

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Englewood Cliffs, NJ: Prentice-Hall.

Elliot, A. J., Eder, A. B., & Harmon-Jones, E. (2013). Approach-avoidance motivation and emotion: Convergence and divergence. *Emotion Review, 5*, 308–311. doi:10.1177/1754073913477517

Fontana, P. C., Cohen, S. D., & Wolvin, A. D. (2015). Understanding listening competency: A systematic review of research studies. *International Journal of Listening, 29*, 148–176. doi:10.1080/10904018.2015.1015226

Ford, W. S. Z., Wolvin, A. D., & Chung, S. (2000). Students' self-perceived listening competencies in the basic speech communication course. *International Journal of Listening, 14*, 1–13. doi:10.1080/10904018.2000.10499032

Gollwitzer, P. M., Fujita, K., & Oettingen, G. (2004). Planning and the implementation of goals. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 211–228). New York: Guilford.

Halone, K. K., Cunconan, T. M., Coakley, C. G., & Wolvin, A. D. (1998). Toward the establishment of general dimensions underlying the listening process. *International Journal of Listening, 12*, 12–28. doi:10.1080/10904018.1998.10499016

Halone, K. K., Wolvin, A. D., & Coakley, C. G. (1997). Accounts of effective listening across the lifespan: Expectations and experiences associated with competent listening practices. *International Journal of Listening, 11*, 15–38. doi:10.1207/s1932586xijl1101_2

Harter, J. K., & Schmidt, F. L. (2008). Conceptual versus empirical distinctions among constructs: Implications for discriminant validity. *Industrial and Organizational Psychology, 1*, 36–39. doi:10.1111/j.1754-9434.2007.00004.x

Hinrichs, J. R. (1964). Communication activity of industrial research personnel. *Personnel Psychology, 17*, 193–204.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465–497). San Diego, CA: Academic Press.

Hoyle, R. H., & Leary, M. R. (2009). Methods of the study of individual differences in social behavior. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 12–23). New York: Guilford.

Iida, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Using diary methods in psychological research. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Volume I. Foundations, planning, measures, and psychometrics* (pp. 277–305). Washington, DC: APA.

Imhof, M. (2004). Who are we as we listen? Individual listening profiles in varying contexts. *International Journal of Listening, 18*, 36–46. doi:10.1080/10904018.2004.10499061

Jefferson, G. (1980). On "trouble-premonitory" response to inquiry. *Sociological Inquiry, 50*, 153–185. doi:10.1111/j.1475-682X.1980.tb00019.x

Jones, S. M. (2011). Supportive listening. *International Journal of Listening, 25*, 85–103. doi :10.1080/10904018.2011.536475

Keaton, S. A., & Bodie, G. D. (2013). The statistical and methodological acuity of scholarship appearing in the *International Journal of Listening* (1987–2011). *International Journal of Listening*, 115–135. doi:10.1080/10904018.2013.813206

Knee, C. R. (1998). Implicit theories of relationships: Assessment and prediction of romantic relationship initiation, coping, and longevity. *Journal of Personality and Social Psychology, 74*, 360–370. doi:10.1037/0022-3514.74.2.360

Kotowski, M. R., Levine, T. R., Baker, C. R., & Bolt, J. M. (2009). A multitrait-multimethod assessment of the Verbal Aggressiveness and Argumentativeness Scales. *Communication Monographs, 76*, 443–462. doi:10.1080/03637750903300247

Krzystofiak, F., Cardy, R., & Newman, J. (1988). Implicit personality and performance appraisal: The influence of trait inferences on evaluations of behavior. *Journal of Applied Psychology, 73*, 515–521. doi:10.1037/0021-9010.73.3.515

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. Reis (Ed.), *New directions for methodology of social and behavioral sciences* (Vol. 15, pp. 41–56). San Francisco: Jossey-Bass.

Lawson, M., & Winkelman, C. (2003). The social desirability factor in the measurement of listening skills: A brief report. *Counseling Psychology Quarterly, 16*, 43–45. doi:10.1080/0951507021000050212

Ledbetter, A. M., & Schrodt, P. (2008). Family communication patterns and cognitive processing: Conversation and conformity orientations as predictors of informational reception apprehension. *Communication Studies, 59*, 388–401. doi:10.1080/10510970802467429

Levine, T. R. (2005). Confirmatory factor analysis and scale validation in communication research. *Communication Research Reports, 22*, 335–338. doi:10.1080/00036810500317730

Levine, T. R., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The desirability of using confirmatory factor analysis on published scales. *Communication Research Reports, 23*, 309–314. doi:10.1080/08824090600962698

McCroskey, J. C. (1997). Self-report measurement. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (2nd ed., pp. 191–216). Cresskill, NJ: Hampton.

Metts, S., Sprecher, S., & Cupach, W. R. (1991). Retrospective self-reports. In B. M. Montgomery & S. Duck (Eds.), *Studying interpersonal interaction* (pp. 162–178). New York: Guilford.

Miell, D. E. (1987). Remembering relationship development: Constructing a context for interaction. In R. Burnett, P. McGee, & D. Clarke (Eds.), *Accounting for relationships* (pp. 60–73). London: Methuen.

Myers, S. (2000). Empathic listening: Reports on the experience of being heard. *Journal of Humanistic Psychology, 40*, 148–173. doi:10.1177/0022167800402004

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. doi:hdl.handle.net/2027.42/92167

O'Keefe, B. J. (1988). The logic of message design: Individual differences in reasoning about communication. *Communication Monographs, 55*, 80–103. doi:10.1080/03637758809376159

Perras, M. T., & Weitzel, A. R. (1981). Measuring daily communication activities. *The Florida Speech Communication Journal, 9*, 19–23.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*, 531–544. doi:10.1177/014920638601200408

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517–529. doi:10.1037//0021-9010.87.3.517

Rankin, P. T. (1926). The measurement of the ability to understand spoken language. *Dissertation Abstracts, 12*(1952), 847.

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–21). New York: Guilford Press.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). Cambridge: Cambridge University Press.

Reis, H. T., Gable, S. L., & Maniachi, M. R. (2014). Methods for studying everyday experiences in its natural context. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 373–403). Cambridge: Cambridge University Press.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods, 12*, 762–800. doi:10.1177/1094428109332834

Rogers, C. R. (1955). *Active listening*. Chicago: The Industrial Relations Center of the University of Chicago.

Schrodt, P., & Wheeless, L. R. (2001). Aggressive communication and informational reception apprehension: The influence of listening anxiety and intellectual inflexibility on trait argumentativeness and verbal aggressiveness. *Communication Quarterly, 49*, 53–69. doi:10.1080/01463370109385614

Schrodt, P., Wheeless, L. R., & Ptacek, K. M. (2000). Informational reception apprehension, educational motivation, and achievement. *Communication Quarterly, 48*, 60–73. doi:10.1080/01463370009385580

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods, 19*, 80–110. doi:10.1177/1094428115598239

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. doi:10.1146/annurev.clinpsy.3.022806.091415

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall, mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 14*, 203–229. doi:10.1111/j.1468-2958.1987.tb00127.x

Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 10*, 379–402. doi:10.1111/j.1468-2958.1987.tb00127.x

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: APA.

Watson, K. W., Barker, L. L., & Weaver, J. B., III. (1995). The listening styles profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening, 9*, 1–13. doi:10.1080/10904018.1995.10499138

Weger, H. W., Jr., Castle, G. R., & Emmett, M. C. (2010). Active listening in peer interviews: The influence of message paraphrasing on perceptions of listening skill. *International Journal of Listening, 24*, 34–49. doi:10.1080/10904010903466311

Weiner, B. (1992). *Human motivation: Metaphors, theories, research*. Newbury Park, CA: Sage.

Weinrauch, J. D., & Swanda, J. R., Jr. (1975). Examining the significance of listening: An exploratory study of contemporary management. *Journal of Business Communication, 13*, 25–32. doi:10.1177/002194367501300103

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*, 339–354. doi:10.1111/j.1467-6494.1991.tb00252.x

Wheeless, L. R. (1975). An investigation of receiver apprehension and social context dimensions of communication apprehension. *The Speech Teacher, 24*, 261–268.

Wheeless, L. R., Preiss, R. W., & Gayle, B. M. (1997). Receiver apprehension, informational receptivity, and cognitive processing. In J. A. Daly, J. C. McCroskey, J. Ayres, T. Hopf, & D. M. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 151–187). Cresskill, NJ: Hampton.

Wolvin, A. D. (2009). Listening, understanding, and misunderstanding. In W. F. Eadie (Ed.), *21st century communication: A reference handbook* (Vol. 1, pp. 137–146). Thousand Oaks, CA: Sage.

Wolvin, A. D., & Coakley, C. G. (1993). A listening taxonomy. In A. D. Wolvin & C. G. Coakley (Eds.), *Perspectives on listening* (pp. 15–22). Norwood, NJ: Ablex.

Wolvin, A. D., & Coakley, C. G. (1994). Listening competency. *Journal of the International Listening Association, 8*, 148–160. doi:10.1080/10904018.1994.10499135

Wolvin, A. D., & Coakley, C. G. (2000). Listening education in the 21st century. *International Journal of Listening, 14*, 143–152. doi:10.1080/10904018.2000.10499040

Worthington, D. L., & Fitch-Hauser, M. E. (2012). *Listening: Processes, functions, and competency*. New York: Routledge.