Chapter 5
Measurement of Empathy in the General Population

If you cannot measure it, you cannot improve it.

(Lord William Thompson Kelvin, 1824–1907)

Abstract

- Some of the instruments that have been developed to measure empathy in children and adults, and used by researchers other than their own authors, are briefly described.
- The three that have been used most often in medical education and health care research are Hogan's Empathy Scale, Mehrabian and Epstein's Emotional Empathy Scale, and Davis's Interpersonal Reactivity Index. These instruments were developed for administration to the general population; therefore their relevance to the context of health professions education and patient care is limited for two reasons. First, as the content of the items in the three instruments implies, none is framed in the context of physician-patient (clinician-client) relationships. Thus, the validity of their use in that context is questionable. Second, the three instruments were not developed specifically to address the cognitively defined concept of empathy, a conceptualization that is more relevant and desirable in the context of patient care.
- The biotechnological advancements in functional brain imaging and the recent discovery of the mirror neuron system have opened up a new window for assessing empathy that is extremely promising.
- Given the findings that empathy tends to erode during medical and other health professions education, and in an era of changes in the health care system that hamper the clinician-patient relationship, a psychometrically sound instrument for measuring empathy in the context of health professions education and patient care is in high demand.

Introduction

In Chap. 1, I indicated that one reason for the dearth of empirical research on empathy in the health professions was the lack of a psychometrically sound instrument that can be used to measure the concept in health professions education and patient care. This chapter briefly describes some of the instruments that have been used most often to measure empathy in the general population and presents sample items enabling us to judge their *face validity* in the context of patient care. Although a detailed analysis of the psychometric properties of these scales can be informative, such a technical discussion is beyond the scope of this book. However, in Chap. 7, I will describe in detail the step-by-step development and psychometric properties of the Jefferson Scale of Empathy, which was specifically designed to measure empathy in medical and other health professions students, practicing physicians, and other health professionals.

In general, an instrument serves not only as a device for measurement but also as the basis of a common language that researchers use to communicate their empirical findings. Therefore, familiarity with the instruments and the scores they generate is necessary to comprehend and compare the results of research. For that purpose, I have selected a few research instruments designed to measure empathy in child and adult populations that are described in the following sections.

Measurement of Empathy in Children and Adolescents

Reflexive or Reactive Crying

Simner (1971) systematically investigated newborn infants' reactive crying and reported that newborns who heard another newborn crying cried significantly more often in response (reflexive crying) than they did to any other nonstartling noise. These findings were later replicated in other studies (Martin & Clark, 1982; Sagi & Hoffman, 1976). It is interesting to note that the newborns did not respond to their own cries (Martin & Clark, 1982) suggesting that infants are capable of distinguishing between self and others early in life (Decety & Jackson, 2004). The reaction of one infant to another infant's crying has been used as an indicator of empathy in infants based on the assumption that an infant crying in response to another infant's distress is a reflection of an empathic response (Eisenberg & Lennon, 1983). Eisenberg (1989) suggested that the capacity to respond to cues of another person's distress in childhood is a primitive precursor of more mature empathic sensibilities that develop later. However, the assumption that reflexive crying in infants is an indicator of empathy needs to be verified empirically in longitudinal studies.

According to Eisenberg and Lennon (1983), no convincing evidence is available to confirm that reflexive crying necessarily implies an empathic response. Martin and Clark (1982) reported that children of both sexes cried more in response to a male newborn's crying than to a female newborn's crying. These reports raise questions about the validity of reflexive crying as an indicator of empathic capacity.

The Picture or Story Methods

One popular method of measuring empathy in young children, developed by Eisenberg and Lennon (1983), has been to expose children to another person's distress by showing them pictures or telling them stories depicting hypothetical situations. The children are subsequently asked to describe their own feelings about the story's protagonist either verbally or by choosing an image from a set of pictures representing a variety of faces exhibiting various expressions, such as a happy or sad face. A match between the child's feelings and the protagonist's feelings is considered to be an indication of empathic understanding. The difficulty of differentiating empathy from sympathy when using picture or story methods of assessment raised concern about the validity of this method. Also, the predictive validity of this method awaits empirical verification.

The Feshbach Affective Situations Test of Empathy

The Feshbach Affective Situations Test of Empathy (FASTE), published by Feshbach and Roe (1968), is a widely used variation of the picture or story method of measuring empathy in children. Children (usually aged 6 or 7 years) are shown cartoons on a series of slides accompanied by hypothetical stories depicting children in different affectively charged conditions (happiness, sadness, fear, and anger). The children are then asked to describe their own feelings and emotions about the picture or story either verbally or by choosing a response from a set of facial expressions depicting different emotions. For example, a theme for happiness is a picture of a birthday party, a theme for sadness is a lost dog, a theme for fear is a frightening dog, and a theme for anger is a false accusation. The child's capacity for empathy is determined by a match between the child's expressed feeling and the theme depicted in the picture or story. The FASTE has been modified to accommodate studies by different researchers (Zhou, Valiente, & Eisenberg, 2003).

Some have criticized the FASTE because of its weak psychometric support, its suggestive test instructions (e.g., instructions designed in a way that elicits the desired behavior), and a lack of clarity in scoring (Eisenberg-Berg & Lennon, 1980; Eisenberg & Lennon, 1983; Hoffman, 1982; Zhou et al., 2003). Concern also has been raised about the confounding effect of the "demand characteristic" in children's responses (Goldstein & Michaels, 1985). This phenomenon makes the respondents modify their responses to what they believe the testing situation demands. The demand characteristic (e.g., a tendency to respond in a certain way that can undermine the validity of the results) is inherent in children's self-reports when an adult constantly asks them about their feelings. Another concern is the confounding effect of the experimenter's gender on the results of the FASTE because research indicated that when the experimenter was a woman, girls scored higher than boys did (Levine & Hoffman, 1975; Roe, 1977).

The Index of Empathy

The self-report Index of Empathy, developed by Bryant (1982), consists of 22 items designed to measure empathy in children and adolescents. The measure is comparable to Mehrabian and Epstein's Emotional Empathy Scale, which was developed to measure empathy in the adult population (this scale will be described later in this chapter). The author of the Index of Empathy indicated that these comparable instruments can be useful for exploring changes in empathy at different ages. A sample item is "I really like to watch people open presents, even when I don't get a present myself." The internal consistency reliability coefficients of this measure were reported to be 0.54 for first graders, 0.68 for fourth graders, and 0.79 for seventh graders (Bryant, 1982; Zhou et al., 2003).

Although the abovementioned methods of measuring empathy in children and adolescents seem to be useful for measuring reactions to affective situations, no convincing evidence is available in support of the instrument's predictive validity as indicators of the capacity for empathy.

Measurement of Empathy in Adults

The Most Frequently Used Instruments

The first three self-report measures of empathy discussed in this section—Hogan's Empathy Scale, Mehrabian and Epstein's Emotional Empathy Scale, and Davis's Interpersonal Reactivity Index—have been the most frequently used instruments in empathy research. Although they were developed for use in the general population, rather than with health professions students and practitioners, they have been used frequently in health care research. These measures are briefly described in the order in which they were originally published. Although other instruments have been designed to measure empathy, they have not received widespread attention. Some of them will be briefly described later in this chapter.

The Empathy Scale

Published by Robert Hogan (1969) and based on his doctoral dissertation at the University of California at Berkeley, the Empathy Scale includes 64 true–false items adopted from the California Psychological Inventory (CPI), the Minnesota Multiphasic Personality Inventory (MMPI), and other tests used at the Institute of Personality Assessment and Research. The scale was developed within the framework of the theory of moral development. A typical item is "I have seen some things so sad that I almost felt like crying."

Evidence in support of the scale's validity was provided by showing that high scorers were more likely than low scorers to be socially acute and sensitive to nuances in interpersonal relationships, and low scorers were more likely to be hostile, cold, and insensitive to the feelings of others (Hogan, 1969). Also, in a group of medical students, Hogan found a significant and positive correlation between scores on this scale and a criterion measure of sociability on the CPI ($r=0.58$) and a significant negative correlation with social introversion on the MMPI ($r=-0.65$) (Hogan, 1969). Factor analysis of the Empathy Scale across different studies resulted in an inconsistent factor structure. For example, Greif and Hogan (1973) reported the following factors: "even-tempered disposition," "social ascendancy," and "humanistic sociopolitical attitudes," and Johnson, Cheek, and Smither (1983) reported "social self-confidence," "even temperedness," "sensitivity," and "nonconformity." These inconsistent findings raised questions about the scale's construct validity. Based on the factor analytic findings, it is suggested that the entire scale may not capture the essence of empathy (Baron-Cohen & Wheelwright, 2004). The scale's reliability has also been questioned (Cross & Sharpley, 1982).

The Emotional Empathy Scale

This instrument was developed by Albert Mehrabian and Norman Epstein (1972) and includes 33 items intended to measure emotional empathy. "It makes me sad to see a lonely stranger in a group" is a typical item. The title of the measure and the contents of the items pertain to susceptibility to emotional contagion (Zhou et al., 2003), indicating that the authors used an affective conceptualization of empathy when developing the scale (Davis, 1994). This conceptualization conflicts with the definition of empathy as a primarily cognitive concept in the context of patient care that was adopted in this book (see Chap. 6).

Items are answered on a 9-point Likert-type scale (Very Strongly Agree=+4, Very Strongly Disagree=−4). The split-half reliability of this scale was reported to be 0.84, and the internal consistency reliability was 0.79 (Zhou et al., 2003). The validity of this scale was determined by using an experimental paradigm similar to Milgram's experiments (1963; 1968) in which high scorers on this scale were less likely than low scores to administer electric shocks to the experimental subjects (Mehrabian & Epstein, 1972) (see Chap. 8 for a description of Milgram's experimental paradigm). On the basis of their subjective view, Mehrabian and Epstein reported that the scale included the following components and identified the items that measured each of these components: extreme emotional responsiveness, appreciation of the feelings of unfamiliar and distant others, tendency to be moved by others' emotional experiences, and tendency to be sympathetic. A study by Dillard and Hunter (1989) failed to support the aforementioned multidimensional components.

Later, Mehrabian, Young, and Sato (1988) changed the scale's name to the Emotional Empathic Tendency Scale. More recently, Mehrabian introduced a new instrument, the Balanced Emotional Empathy Scale (BEES), to measure vicarious

empathy (Mehrabian, 1996) which is similar in content to the Emotional Empathy Scale. It contains 30 items, each answered on a 9-point Likert scale (4 = Very Strongly Agree, −4 = Very Strongly Disagree). A sample item is "Unhappy movie endings haunt me for hours afterward." Information about the scale is posted on Mehrabian's personal website, and to my knowledge no empirical study has been published to specifically address psychometrics of the BEES. In a study of empathy and aggression, a Cronbach's alpha coefficient of 0.87 was reported for this scale (Mehrabian, 1997).

The Interpersonal Reactivity Index

As part of his doctoral dissertation at the University of Texas at Austin, Mark Davis developed the Interpersonal Reactivity Index (IRI) (Davis, 1983) to measure individual differences in empathy. The instrument includes 28 items tapping four components of empathy in the cognitive and emotional domains. These four components are reflected in four subscales (Perspective Taking, Empathic Concern, Fantasy, and Personal Distress), each of which includes seven items answered on a 5-point scale ranging from 0 (Does not describe me well) to 4 (Describes me very well). These components were originally determined by subjective judgment without statistical support. However, confirmatory factor analysis provided mixed results concerning the existence of the four subscales (Cliffordson, 2002; Litvack-Miller, McDougall, & Romney, 1997).

The Perspective Taking subscale measures the tendency to adopt the views of others spontaneously. "I sometimes try to understand my friends better by imagining how things look from their perspective" is a typical item. The Empathic Concern subscale measures a tendency to experience the feelings of others and to feel sympathy and compassion for unfortunate people. A typical item is "I often have tender, concerned feelings for people less fortunate than me." The Fantasy subscale measures a tendency to imagine oneself in a fictional situation. A typical item is "After seeing a play or movie, I have felt as though I were one of the characters." The Personal Distress subscale taps a tendency to experience distress in others. "When I see someone who badly needs help in an emergency, I go to pieces" is a representative item. According to Davis, the Perspective Taking subscale is more likely to measure cognitive empathy, whereas the other three subscales are more likely to measure emotional empathy.

The internal consistency reliability coefficients ranged from 0.71 to 0.77 for the four subscales, and their test–retest reliabilities ranged from 0.62 to 0.71 (Davis, 1983). The test–retest reliabilities in an adolescent sample over a 2-year period ranged from 0.50 to 0.62 (Davis & Franzoi, 1991; Zhou et al., 2003). In correlating the IRI subscale scores with scores on Hogan's Empathy Scale, the highest positive correlation was found for the Perspective Taking subscale ($r = 0.40$) and the highest negative correlation was found for the Personal Distress subscale ($r = −0.33$) (Davis, 1983). The Perspective Taking subscale of the IRI yielded the lowest correlation with the scores of Mehrabian and Epstein's Emotional Empathy Scale ($r = 0.24$),

and the Fantasy and Empathic Concern subscales yielded the highest correlations (0.52 and 0.60, respectively) (Davis, 1983). This pattern of correlations confirms Davis's claim about the cognitive nature of the Perspective Taking subscale.

In 1994, Davis stated that convincing evidence existed in support of some psychometric aspects of the IRI, although no satisfactory statistical evidence has been presented to confirm the stability of the four components of the index. In a study with physicians and undergraduate psychology students, Yarnold, Bryant, Nightingale, and Martin (1996) discovered an additional component called "involvement" in their statistical analysis of the IRI. The findings that scores on the Personal Distress subscale of the IRI were negatively correlated with scores on the Perspective Taking subscale raise a serious question about the validity of scoring the IRI by summing up the scores of all its subscales, including the Personal Distress subscale. According to D'Orazio (2004), because of the negative correlation between the Personal Distress and Perspective Taking subscales and because high scores on Personal Distress are associated with dysfunctional interpersonal relationships, summing up the scores of all four subscales of the IRI would not be meaningful.

Other Instruments

Several other instruments for measuring empathy in the adult population are described here in chronological order. Kerr developed a test of empathy with the intention of measuring respondents' ability to "anticipate" certain typical reactions, feelings, and behavior of other people (Kerr, 1947). The test consists of three sections which require respondents to rank the popularity of 15 types of music, the national circulation of 15 magazines, and the prevalence of 10 types of annoyances for a particular group of people (Chlopan, McCain, Carbonell, & Hagen, 1985). The respondent's rankings are compared to the empirical data to assess the accuracy of the respondent's rankings. This test seems to be a measure of general information, rather than a measure of empathy. Nevertheless, Kerr and Speroff (1954) claimed that the test was an indicator of empathic understanding and that it could predict a person's popularity, feelings for others, leadership, and sales records.

A measure of insight and empathy was introduced by Dymond (1949, 1950). This measure was based on the conceptualization of empathy as the imaginative transposing of oneself into another person's thinking, feeling, and acting. In Dymond's Rating Test (of empathic ability), respondents rate themselves and one another on a 5-point scale on six attributes such as "superior–inferior," "friendly–unfriendly," "leader–follower," "self-confidence," "selfish–unselfish," and "sense of humor." The concordance between individual's ratings of himself or herself and the individual's predictions of how others would rate him or her was considered as a measure of empathic ability. High scorers on the Dymond's Rating Test were classified as empathizers by analyses of their responses to the Thematic Apperception Test (TAT) (Dymond, 1949). Although no satisfactory evidence is available to confirm the instrument's validity as a measure of empathy, some preliminary data on its psychometric characteristics were

presented by Chlopan et al. (1985). However, those investigators raised concerns about the measure's lack of easy administration and scoring procedures.

Barrett-Lennard (1962) developed an instrument called the Relationship Inventory, which was designed to investigate changes in the clinician–client relationship in the psychotherapeutic context. The instrument can be completed by either the clinician or the client. The original inventory included 92 items. However, one revised version consists of 64 items divided into four subjectively determined subtests of interpersonal relationships: (1) Empathic Understanding, described as the extent to which one person is conscious of the awareness of another person; (2) Level of Regard, the affective aspect of one person's response to another; (3) Unconditionality of Regard, the degree of constancy of regard one person feels for another person; and (4) Congruence, the degree to which one person is functionally integrated in the context of his or her relationship with another person (Barrett-Lennard, 1986). The "Willingness To Be Known" subtest included in the original version of the Relationship Inventory was defined as the degree to which a person wants to be known as a person by another person. This subtest was dropped in the revised version because of its nonsignificant predictive validity concerning therapeutic outcomes (Barrett-Lennard, 1986). Subsequent versions of the Relationship Inventory have been developed for use in nonclinical situations involving family, friendship, coworker, and teacher–pupil relationships (Bennett, 1995).

A 16-item subtest of this instrument called Empathic Understanding contains such items as "He [clinician/client] understands me." Items are answered on a Likert-type scale ranging from −3 ("No," as strongly felt disagreement) to +3 ("Yes," as strongly felt agreement). A negligible clinician–client correlation of 0.09 was reported for the Empathic Understanding subtest (Barrett-Lennard, 1962).

Truax and Carkhuff (1967) developed the 141-item Relationship Questionnaire to measure clients' perceptions of psychologists or counselors in psychological counseling and psychotherapy. Forty-six of the 141 items of the Relationship Questionnaire form a subscale called the Accurate Empathy Scale, which consists of such items as "He sometimes completely understands me so that he knows what I am feeling even when I am hiding my feelings." A number of questions have been raised about the validity, reliability, and score stability of the Accurate Empathy Scale (Beutler, Johnson, Neville, & Workman, 1973; Blass & Hech, 1975; Chinsky & Rappaport, 1970).

Carkhuff (1969) developed the Empathic Understanding in Interpersonal Processes Scale. This single-item instrument gives clinicians an overall empathy score based on five levels of empathic behavior, as judged by observers. Clinicians who score at Level 1 are judged as unable to express any awareness of even the most obvious of a client's feelings, whereas those who score at Level 5 are judged to be fully aware of and able to respond accurately to all of the client's feelings. Because an observer rates clinicians' empathic global behavior on a single item, the validity of this instrument is questionable (LaMonica, 1981).

The Fantasy-Empathy (F-E) Scale developed by Stotland, Mathews, Sherman, Hansson, and Richardson (1978) measures the tendency to respond emotionally to situations. The scale contains three items answered on a 5-point scale: for example, "When I watch a good movie, I can very easily put myself in the place of a leading

character." Some psychometric data on this brief scale have been reported (Stotland, 1978). For instance, a correlation of 0.44 was reported between scores of the F-E Scale and Mehrabian and Epstein's Emotional Empathy Scale (Williams, 1989).

Layton (1979) developed the Empathy Test, a two-part 48-item instrument, as part of a research project designed to teach empathy to nursing students. The purpose of this measure was to evaluate whether empathy can be learned by observing models of empathic behavior. Each part of the Empathy Test consists of 12 true–false items and 12 multiple-choice items. According to Layton's reports, the reliability coefficients for the measure are unacceptably low (in the 0.20s), and no significant correlations were found between this measure and the Empathic Understanding subtest of Barrett-Lennard's Relationship Inventory and Carkhuff's Empathic Understanding in Interpersonal Processes Scale (Carkhuff, 1969).

Another instrument for measuring empathy is the Empathy Construct Rating Scale developed by LaMonica (1981). The instrument consists of 84 items about the respondent's feelings or actions toward another person, answered on a 6-point Likert-type scale (−3, Extremely Unlike; +3, Extremely Like). A typical item is "Seems to understand another person's state of being." The bipolar grand factor of this scale includes the notion of "well-developed empathy" (e.g., "Shows consideration for a person's feelings and reactions") at one pole and "lack of empathy" (e.g., "Does not listen to what the other person is saying") at the opposite pole.

A 15-item unidimensional instrument (The Emotional Contagion Scale) was developed by Doherty (1997) to measure emotional empathy. Each item is answered on a 5-point Likert-type scale. A sample item is "I cry at sad movies." A Cronbach's alpha coefficient of 0.90 is reported for the scale. Higher correlation was found between scores of this scale and those of the Empathic Concern scale ($r = 0.37$) than scores of the Perspective Taking scale of the IRI ($r = 0.14$).

A measuring instrument—Empathy Quotient (EQ)—was developed in England by Baron-Cohen and Wheelwright (2004) that contains 40 empathy items plus 20 filler items to distract the participants from relentless focus on empathy (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004). Each item is answered on a 4-point Likert-type scale from Strongly Agree to Strongly Disagree. Although the authors claim that the EQ was explicitly designed to have clinical applications, the contents of most of the items do not support such an application. Sample items are "I really enjoy caring for other people" and "I tend to get emotionally involved with a friend's problems." A test–retest reliability of 0.83 is reported for the EQ. Three factors, Cognitive Empathy, Emotional Reactivity, and Social Skills, emerged from factor analyses of the EQ. With the exception of the Cognitive Empathy factor, which was not correlated with any subscales of the IRI, the EQ yielded moderate correlations with the Empathic Concern and Perspective Taking subscales of the IRI and a negligible negative correlation with the Personal Distress subscale (Lawrence et al., 2004).

Another empathy measuring instrument for the general population is the Toronto Empathy Questionnaire (TEQ) for measuring emotional empathy (Spreng, McKinnon, Mar & Levine et al, 2009). This instrument contains 16 questions. A sample item is "I become irritated when someone cries." Some psychometric data exist in support of the validity and reliability of the TEQ in a study by its authors (Spreng et al., 2009).

Mercer, Maxwell, Heaney, and Watt (2004) and Mercer, McConnachie, Maxwell, Heaney, and Watt (2005) developed the Consultation and Relational Empathy (CARE) instrument for administration to patients to assess their doctors' or health care providers' empathic engagement in clinical encounters. This instrument includes ten items; each is answered on a 5-point Likert scale (1 = Poor, 5 = Excellent). A sample item is "How was the doctor at being interested in you as a whole person." Data in support of validity of the instrument and a Cronbach's alpha coefficient of 0.92 have been reported by the test authors (Mercer, Maxwell, Heaney, and Watt (2004).

There are a few review articles about empathy measuring instruments. For example, Yu and Kirk (2009) identified 20 empathy measures used in nursing research, and concluded that none of the reviewed measures was psychometrically robust. Hemmerdinger, Stoddart, and Lilford (2007) reported that based on their systematic review of the literature, 36 empathy measuring instruments were identified, but only eight demonstrated evidence in support of their validity, internal consistency, and reliability. There are other instruments, claimed by their authors as measures of empathy; however, either no convincing evidence has been presented to support their psychometrics, or they have not been used by other researchers except their own authors.

Physiological and Neurological Indicators of Empathy

Some social psychologists have studied empathy by using physiological measures, such as heart rate, skin conductance, palmar sweating, and vasoconstriction, as indicators of understanding other people's distress (Goldstein & Michaels, 1985; Stotland et al., 1978). Although most of these physiological measures are likely to be free of a social desirability response bias, they seem to be indicative of a person's emotional reaction to another person's distress. Such physiological reactions are more likely to be akin to sympathy than to empathy. Correspondingly, they may not be appropriate for the measurement of cognitively defined empathy in patient care.

Recently, functional brain-imaging methods (e.g., fMRIs and PET scans) have been used as indicators of brain activity in individuals experiencing empathy (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Wicker et al., 2003). In addition to advancements in functional brain imaging, the discovery of the mirror neuron system activated by observing another person in pain or performing an act (see Chap. 13) is, I believe, the beginning of a promising approach to quantifying neurophysiological manifestations of empathy in future research (see Chap. 13 for more detailed discussion).

Relationships Among Measures of Empathy

The results of studies attempting to determine correlations among different measures of empathy have not been encouraging. For example, Jarski, Gjerde, Bratton, Brown, and Matthes (1985) tested a group of medical students and found no

significant correlations among the Empathy Scale (Hogan, 1969), the Empathic Understanding subtest of the Relationship Inventory (Barrett-Lennard, 1962), or the Empathic Understanding in Interpersonal Processes Scale (Carkhuff, 1969).

Another study with registered nurses examined correlations among four measures of empathy (Layton & Wykle, 1990). The results showed that Carkhuff's Empathic Understanding in Interpersonal Processes Scale was moderately correlated ($r=0.25$) with Layton's Empathy Test but was not correlated with the Empathic Understanding subtest of Barrett-Lennard's Relationship Inventory. In addition, LaMonica's Empathy Construct Rating Scale was not correlated with Layton's Empathy Test but was moderately correlated ($r=0.37$) with Carkhuff's Empathic Understanding in Interpersonal Processes Scale and highly correlated ($r=0.78$) with the Barrett-Lennard's Empathic Understanding subtest of the Relationship Inventory.

In a review article, Chlopan et al. (1985) reported the findings of studies on the validity and reliability of several measures of empathy, including Mehrabian and Epstein's Emotional Empathy Scale and Hogan's Empathy Scale. They argued that both of these scales seem to measure two different aspects of empathy. As its name indicates, the Emotional Empathy Scale is more likely to measure the affective aspects of empathy, or general emotional arousability (Mehrabian et al., 1988), whereas the Empathy Scale is more likely to measure role-taking ability, a cognitive aspect of empathy. Chlopan and colleagues also reported that the subscales of the IRI seem to tap both the emotional (e.g., Personal Distress subscale) and the cognitive (Perspective Taking subscale) aspects of empathy.

The intercorrelations among these empathy measures are often weak and inconsistent and, in most cases, nonsignificant or negligible (Bohart, Elliot, Greenberg, & Watson, 2002; Gladstein & Associates, 1987). One reason for these inconsistent findings is that different instruments tap different aspects of empathy based on different definitions of the concept. Although these instruments can have potential value in particular situations, none can be recommended as the best for all patient-care situations (Bennett, 1995). With the exception of the Perspective Taking subscale of the IRI, the contents of the other instruments described in this chapter do not reflect the cognitive conceptualization of empathy adopted in this book (see Chap. 6). Thus, their face validity (and content validity) would be questionable when empathy is conceptualized as a predominantly cognitive attribute in the context of patient care advocated in this book.

A Need for an Instrument Specifically Designed to Measure Empathy in Patient Care

A measure that assesses empathy in patient care—particularly in medical and surgical treatment—needs to be more specific than the instruments I have discussed in this chapter so far. Because of the findings on the decline of empathy during health professions education (see Appendix A), and changes evolving in the market-driven

health care systems that hamper clinician-patient relationships, the empirical study of empathy in health care education and practice is both important and timely. Among prerequisites to empirical research on empathy in the health professions education and the practice of patient care are (1) an operational definition of the concept (see Chaps. 1 and 6), and (2) a psychometrically sound instrument for quantifying the concept in the context of health professions education and patient care. In 2000, in response to a need for a psychometrically sound instrument to measure empathy in the context of patient care, our research team in the Center for Research in Medical Education and Health Care at Jefferson (currently Sidney Kimmel) Medical College developed an instrument specifically designed to measure empathy among students and practitioners in the health care professions. This scale will be described in detail in Chap. 7.

Recapitulation

Several instruments exist that claim to measure empathy in children and adults. The three frequently used instruments intended to measure empathy in adults—Hogan's Empathy Scale, Mehrabian and Epstein's Emotional Empathy Scale, and Davis's IRI—were developed for administration to the general population. The examination of their contents suggests that they do not tap the essence of empathy in the context of health professions education and patient care. In other words, their face and content validities in the context of health professions education and patient care are questionable. Recently, functional brain imaging technology that has been used to address brain activities in interpersonal relationships has emerged as a promising path for measuring empathic engagement. I suspect that the findings of most of the studies in which the instruments described in this chapter were used are questionable in addressing empathy issues in the context of patient care. The reason is that the content of the bulk of the items in the self-reported instruments, described in this chapter, taps on feeling the pain and suffering of others (described as emotional or affective empathy, synonymous to sympathy and arousability) rather than empathic understanding (e.g., cognitive empathy) which has a different consequence in patient care (see Chaps. 1 and 6). Thus, there was a need for an instrument specifically developed to measure empathy in the context of patient care which will be described in Chap. 7.