Several different sets of criteria for evaluation of theories have been published (Barnum, 1988; Duffy & Muhlenkamp, 1974; George, 2002; Marriner-Tomey & Alligood, 2002; Meleis, 1997; Parse, 1987). Just one set of criteria, however, differentiates between grand theories and middle-range theories (Fawcett, 2000, 2005). Although none of the authors of those publications have indicated that the source of the data for a theory influences the selection of evaluation criteria, a recent conversation with a colleague raised the question of whether theories grounded in qualitative data should be evaluated using criteria that differ from those used to evaluate theories grounded in quantitative data. That conversation resulted in an invitation to Rosemarie Rizzo Parse to engage in a dialogue about what criteria are appropriate for evaluating grand theories and middle-range theories and whether those criteria can be applied to theories regardless of the type of data (qualitative or quantitative) in which a theory is grounded.

Consensus exists that theories are made up of ideas called concepts and statements about the concepts, called propositions (King & Fawcett, 1997). Consensus also exists that components of nursing knowledge, including theories, vary in levels of abstraction (King & Fawcett). I regard grand theories as more abstract than middle-range theories but less abstract than conceptual models (Fawcett, 2005). Accordingly, my framework includes some differences in the evaluative criteria for grand theories and middle-range theories. I do not, however, differentiate criteria for evaluation of either grand or middle-range theories based on the type of data (qualitative or quantitative) used to develop the theory.

I developed a framework for both the analysis and evaluation of nursing theories several years ago and refined the framework twice (Fawcett, 1993, 2000, 2005). Analysis involves objective and nonjudgmental descriptions of theories, whereas evaluation involves judgments about the extent to which nursing theories meet certain criteria. For the purposes of this dialogue, my comments are limited to the criteria for evaluation of grand theories and middle-range theories. Those criteria are significance, internal consistency, parsimony, testability, empirical adequacy, and pragmatic adequacy.

The criterion of significance focuses on the context of the theory. That criterion requires justification of the importance of the theory to the discipline of nursing and is met when the metaparadigmatic, philosophical, and conceptual origins of the theory are explicit, when antecedent nursing and adjunctive knowledge is cited (Levine, 1988), and when the special contributions made by the theory are identified. The four questions to be asked when evaluating the significance of a theory, which are applicable to both grand and middle-range theories, are listed in Table 1.

The criterion of internal consistency focuses on both the context and the content of the theory. That criterion requires all elements of the theorist’s work, including the philosophical claims, conceptual model, and theory concepts and propositions, to be congruent. The internal consistency criterion also requires the concepts of the theory to reflect semantic clarity and semantic consistency. The semantic clarity re-

Keywords: evaluation criteria, grand theory, middle-range theory, nursing theories
requirement is more likely to be met when a theoretical definition is given for each concept than when no explicit definitions are given. The semantic consistency requirement is met when the same term and the same definition are used for each concept in all of the author’s discussions about the theory. Semantic inconsistency occurs when different terms are used for a concept or different meanings are attached to the same concept. In addition, the internal consistency criterion requires that propositions reflect structural consistency, which means that the linkages between concepts are specified and that no contradictions in relational propositions are evident. The three questions to be asked when evaluating the internal consistency of a theory, which is applicable to both grand and middle-range theories, are listed in Table 1.

The criterion of parsimony focuses on the content of the theory. Parsimony requires a theory to be stated in the most economical way possible without oversimplifying the phenomena of interest. This means that the fewer the concepts and propositions needed to fully explicate the phenomena of interest, the better. The parsimony criterion is met when the most parsimonious statements clarify rather than obscure the phenomena of interest. The question to be asked when evaluating the parsimony of a theory, which are applicable to both grand and middle-range theories, is given in Table 1.

The criterion of testability also focuses on the content of the theory. That criterion frequently is regarded as the major characteristic of a scientifically useful theory. Marx (1976) declared, “If there is no way of testing a theory it is scientifi-

---

### Table 1

**Fawcett’s Criteria for Evaluation of Nursing Theories and Pertinent Questions**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Pertinent Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance</td>
<td>Are the metaparadigm concepts and propositions addressed by the theory explicit? Are the philosophical claims on which the theory is based explicit? Are the concepts model from which the theory was derived explicit? Are the authors of antecedent knowledge from nursing and adjunctive disciplines acknowledged and are bibliographical citations given?</td>
</tr>
<tr>
<td>Internal Consistency</td>
<td>Are the context (philosophical claims and conceptual model) and the content (concepts and propositions) of the theory congruent? Do the concepts reflect semantic clarity and semantic consistency? Do the propositions reflect structural consistency?</td>
</tr>
<tr>
<td>Parsimony</td>
<td>Is the theory content stated clearly and concisely?</td>
</tr>
<tr>
<td>Testability: Grand Theories</td>
<td>Is the research methodology qualitative and inductive? Is the research methodology congruent with the philosophical claims and content of the grand theory? Will the data obtained from use of the research methodology represent sufficiently in-depth descriptions of one or more personal experience(s) to capture the essence of the grand theory?</td>
</tr>
<tr>
<td>Testability: Middle-Range Theories</td>
<td>Does the research methodology reflect the middle-range theory? Are the middle-range theory concepts observable through instruments that are appropriate empirical indicators of those concepts? Do the data analysis techniques permit measurement of the middle-range theory propositions?</td>
</tr>
<tr>
<td>Empirical Adequacy: Grand Theories</td>
<td>Are the findings from studies of descriptions of personal experiences congruent with the concepts and propositions of the grand theory?</td>
</tr>
<tr>
<td>Empirical Adequacy: Middle-Range Theories</td>
<td>Are theoretical assertions congruent with empirical evidence?</td>
</tr>
<tr>
<td>Pragmatic Adequacy</td>
<td>Are education and special skill training required before application of the theory in nursing practice? Has the theory been applied in the real world of nursing practice? Does the practitioner have the legal ability to implement and measure the effectiveness of theory-based nursing actions? Are the theory-based nursing actions compatible with expectations for nursing practice? Do the theory-based nursing actions lead to favorable outcomes? Is the application of theory-based nursing actions designed so that comparisons can be made between outcomes of use of the theory and outcomes in the same situation when the theory was not used? Are outcomes measured in terms of the problem-solving effectiveness of the theory?</td>
</tr>
</tbody>
</table>

are asked when evaluating the testability of a middle-range requirements identified by Silva (1986) and Fawcett (1999), techniques are available to measure the assertions made by been developed to observe the theory concepts and statistical when specific instruments or experimental protocols have criterion of testability for middle-range theories, then, is met when evaluating the empirical adequacy of a grand theory involves determining the middle-range theory-generating capacity of a grand theory. The criterion of testability is met when the grand theory has led to the generation of one or more middle-range theories. Three questions, which were adapted from requirements proposed by Silva and Sorrell (1992), are asked when evaluating the testability of a grand theory (Table 1).

The criterion of testability for middle-range theories means that their concepts can have operational definitions and their propositions are amenable to direct empirical testing. Consequently, an approach called traditional empiricism is used to evaluate the testability of middle-range theories. That approach requires the concepts of a middle-range theory to be observable and the propositions to be measurable. Concepts are empirically observable when operational definitions identify the empirical indicators that are used to measure the concepts. Propositions are measurable when empirical indicators can be substituted for concept names in each proposition and when statistical procedures can provide evidence regarding the assertions made. The criterion of testability for middle-range theories, then, is met when specific instruments or experimental protocols have been developed to observe the theory concepts and statistical techniques are available to measure the assertions made by the propositions. Three questions, which were adapted from requirements identified by Silva (1986) and Fawcett (1999), are asked when evaluating the testability of a middle-range theory (Table 1).

The criterion of empirical adequacy requires the assertions made by the theory to be congruent with empirical evidence. The extent to which a theory meets that criterion is determined by means of a systematic review of the findings of all studies that have been guided by the theory. The logic of scientific inference dictates that if the empirical data conform to the theoretical assertions, it may be appropriate to tentatively accept the assertions as reasonable or adequate. Conversely, if the empirical data do not conform to the assertions, it is appropriate to conclude that the assertions are false. Evaluation of the empirical adequacy of a theory should take into consideration the potential for circular reasoning. More specifically, if data always are interpreted in light of a particular theory, it may be difficult to see results that are not in keeping with that theory. Indeed, if researchers constantly uncover, describe, and interpret data through the lens of a particular theory, the outcome may be limited to expansion of that theory and that theory alone (Ray, 1990). Therefore, unless alternative theories are considered when interpreting data or the data are critically examined for both their fit and nonfit with the theory, circular reasoning will occur and the theory will be uncritically perpetuated. Circular reasoning can be avoided if the data are carefully examined to determine the extent of their congruence with the concepts and propositions of the theory, as well as from the perspective of alternative theories (Platt, 1964). In other words, evaluation of a theory always should take alternative theories into account when interpreting data collected within the context of the theory in question.

It is unlikely that any one test of a theory will provide the definitive evidence needed to establish its empirical adequacy. Thus decisions about empirical adequacy should take the findings of all related studies into account. Meta-analysis and other formal procedures can be used to integrate the results of related studies. It is important to point out that a theory should not be regarded as the truth or an ideology that cannot be modified. Indeed, no theory should be considered final or absolute, because it is always possible that subsequent studies will yield different findings or that other theories will provide a better fit with the data. Thus the aim of evaluation of empirical adequacy is to determine the degree of confidence warranted by the best empirical evidence, rather than to determine the absolute truth of the theory. The outcome of evaluation of empirical adequacy is a judgment regarding the need to modify, refine, or discard one or more concepts or propositions of the theory.

The extent to which a grand theory meets the criterion of empirical adequacy is determined by a continuation of the description of personal experiences approach discussed earlier in the section on testability of grand theories. The data used to determine the empirical adequacy of a grand theory may come from multiple personal experiences of an individual or similar personal experiences of several individuals. The extent to which a middle-range theory meets the criterion of empirical adequacy is determined by a continuation of the traditional empirical approach discussed earlier in the section on testability of middle-range theories. The questions to be asked when evaluating the empirical adequacy of grand and middle-range theories are listed in Table 1.

The criterion of pragmatic adequacy focuses on the utility of the theory for nursing practice. The extent to which a grand theory or a middle-range theory meets this criterion is determined by reviewing all descriptions of the use of the theory in practice. The pragmatic adequacy criterion requires that
nurses have a full understanding of the content of the theory, as well as the interpersonal and psychomotor skills necessary to apply it (Magee, 1994). Although that may seem obvious, it is important to acknowledge the need for education and special skill training before theory application.

The pragmatic adequacy criterion also requires that the theory actually is used in the real world of nursing practice (Chinn & Kramer, 1995). In addition, the pragmatic adequacy criterion requires that the application of the theory-based nursing actions is generally feasible (Magee, 1994). Feasibility is determined by an evaluation of the availability of the human and material resources needed to establish the theory-based nursing actions as customary practice, including the time needed to learn and implement the protocols for nursing actions; the number, type, and expertise of personnel required for their implementation; and the cost of in-service education, salaries, equipment, and protocol-testing procedures. Moreover, the willingness of those who control financial resources to pay for the theory-based nursing actions, such as healthcare system administrators and third-party payers, must be determined. In sum, the nurse must be in a setting that is conducive to application of the theory and have the time and training necessary to apply it.

Furthermore, the pragmatic adequacy criterion requires the nurse to have the legal ability to control the application and to measure the effectiveness of the theory-based nursing actions. Such control may be problematic in that nurses are not always able to carry out legally sanctioned responsibilities because of resistance from others. Sources of resistance against implementation of theory-based nursing actions include attempts by physicians and healthcare system administrators to control nursing practice, financial barriers imposed by healthcare institutions and third-party payers, and skepticism by other health professionals about the ability of nurses to carry out the proposed actions (Funk, Tornquist, & Champagne, 1995). The cooperation and collaboration of others may, therefore, have to be secured.

Moreover, the pragmatic adequacy criterion requires that theory-based nursing actions be compatible with expectations for practice (Magee, 1994). Compatibility should be evaluated in relation to expectations held by the public and the healthcare system. If the actions do not meet existing expectations, they should be abandoned or people should be helped to develop new expectations. Johnson (1974) commented, “Current [nursing] practice is not entirely what it might become and [thus people] might come to expect a different form of practice, given the opportunity to experience it” (p. 376).

The pragmatic adequacy criterion also requires the theory-based nursing actions to be socially meaningful by leading to favorable outcomes for those who participate in the actions. Examples of favorable outcomes include a reduction in complications, improvement in health conditions, and increased satisfaction with the theory-based actions on the part of all who participate.

The outcomes of theory-based nursing actions are further judged by use of what Silva and Sorrell (1992) called the problem-solving approach. That approach emphasizes the problem-solving effectiveness of a theory and seeks to determine “whether what is purported or experienced accomplishes its purpose” (Silva & Sorrell, 1992, p. 19). The problem-solving approach is based on the position that theories are developed “to solve human and technical problems and to improve practice” (Kerlinger, 1979, p. 280). It requires deliberative application of a theory. Chinn and Kramer (1995) explained that the application “involves using research methods to demonstrate how a theory affects nursing practice and places the theory within the context of practice to ensure that it serves the goals of the profession . . . [and] provides evidence of the theory’s usefulness in ensuring quality of care” (p. 164). The problem-solving approach can be used with all types of theories but is most effective when applied to middle-range predictive theories. In that case, the application seeks to determine the effects of interventions specified in middle-range predictive theories on the health conditions of the human beings who participate in the interventions (Hegyvary, 1992). The eight questions to be asked when evaluating pragmatic adequacy are given in Table 1. Two last two questions, which were adapted from requirements identified by Silva and Sorrell (1992), are asked when evaluating problem-solving effectiveness of nursing theories.

References
I am pleased to share with Dr. Fawcett my criteria for evaluation of theory and a comparison of our work. First, I should clarify my position on middle-range theory, since it is in direct contrast to Dr. Fawcett’s. The term middle-range theory is ubiquitous in the nursing literature without any substantive definition. Cody (1999) said that there is "a lack of clarity as to what constitutes middle-range" (p. 10). He raised this question: Are those “working in the middle range on myriad topics . . . really developing nursing science or merely elaborating the vast patchwork quilt of applied-science nursing”? (Cody, 1999, p. 11)? Dr. Fawcett and I generally agree that nursing is a basic science with its own unique body of knowledge and that theory is defined as a set of concepts combined uniquely and written at an abstract level to describe, explain, or predict phenomena (Parse, 1997). My view is that propositional statements written at lower levels of abstraction, often called middle-range theories, are really hypotheses that can be tested only through quantitative research methods using appropriate instrumentation (Parse, 2000). Thus, my comments here do not address middle-range theories and are relevant only to theory as described above.

I set forth criteria for the evaluation of nursing theory in my 1987 book, Nursing Science: Major Paradigms, Theories, and Critiques. At that time, I developed a design for critical appraisal appropriate for all frameworks and theories, no matter how they are constructed. I have modified these ideas considerably to be consistent with my current thinking, but I still believe in the basic premise that criteria for evaluation of theory should be broad enough to accommodate all perspectives in the discipline. The two major areas of critical appraisal in my design are structure and process (Parse, 1987). Questions applicable to both areas are listed in Table 1.

Structure criteria refer to the physiognomy of the theory, that is, the historical evolution, foundational elements, and relational statements. Historical evolution refers to the details of the development of the theory including the philosophical and theoretical antecedents and the changes in the theory over time. Foundational elements refer to the philosophical assumptions underpinning the theory and the major concepts of the theory. These are written at an abstract level and are the theorist’s beliefs about the phenomenon of concern to the discipline, the human-universe-health process. Relational statements refer to the principles that are created from the unique weaving of the concepts into descriptions of the human-universe-health process, and they also are written at an abstract level.

The process criteria encompass correspondence, coherence, and pragmatics (Parse, 1987). Correspondence refers to semantic integrity and simplicity. Semantic integrity is recognized by the consistency of meanings among the terms used to explain the human-universe-health process in the philosophical assumptions and in the definitions of the concepts and principles. Two aspects of semantic integrity are substance and clarity (Parse, 1987). Substance refers to the durability of the meaning assigned to terms, the breadth of the descriptions, and the consistency in levels of discourse within and among the assumptions, concepts, and principles. Clarity refers to the distinctness and mutually exclusive nature of the definitions. Simplicity is recognized by the uncluttered abstract descriptions and economy of words used to explain the theory.