Measuring the Effectiveness of Faculty Mentoring Relationships
Ronald A. Berk, PhD, Janet Berg, MS, RN, Rosemary Mortimer, MS, MSEd, RN, Benita Walton-Moss, DNS, RN, and Theresa P. Yeo, MSN, MPH, RN

Abstract

“Mentor” is a term widely used in academic medicine but for which there is no consensus on an operational definition. Further, criteria are rarely reported for evaluating the effectiveness of mentoring. This article presents the work of an Ad Hoc Faculty Mentoring Committee whose tasks were to define “mentorship,” specify concrete characteristics and responsibilities of mentors that are measurable, and develop new tools to evaluate the effectiveness of the mentoring relationship. The committee developed two tools: the Mentorship Profile Questionnaire, which describes the characteristics and outcome measures of the mentoring relationship from the perspective of the mentee, and the Mentorship Effectiveness Scale, a 12-item six-point agree–disagree-format Likert-type rating scale, which evaluates 12 behavioral characteristics of the mentor. These instruments are explained and copies are provided. Psychometric issues, including the importance of content-related validity evidence, response bias due to acquiescence and halo effects, and limitations on collecting reliability evidence, are examined in the context of the mentor–mentee relationship. Directions for future research are suggested.


O
ver the past 25 years, there has been a lack of clarity about the characteristics and outcomes of mentoring relationships, despite a growing body of research. “Mentor” has taken on numerous meanings and has been applied in a variety of corporate and educational contexts since its origin about 2,600 years ago, give or take a month or two: In The Odyssey, the ancient Greek poet Homer recounts the saga of Odysseus, the Greek king and warrior. When he knew he would be away from home for many years, he chose a trusted friend, Mentor, to educate, tutor, protect, and guide his son. Since the mid-1970s, more than 20 definitions of mentoring or mentors have appeared in the literature. These definitions are extremely diverse, plus there is no professional consensus on any “acceptable” definition. Wrightsman’s observation over 20 years ago still seems apropos today: “There is a false sense of consensus, because at a superficial level everyone “knows” what mentoring is. But closer examination indicates wide variation in operational definitions, leading to conclusions that are limited to the use of particular procedures.”

Another strategy to provide some meaning to the construct of mentorship has been to identify the basic elements or functions of the mentoring relationship. Jackson distilled five elements in the mentoring relationship on which there is general agreement. A mentoring relationship (1) focuses on achievement or acquisition of knowledge; (2) consists of three components: emotional and psychological support, direct assistance with career and professional development, and role modeling; (3) is reciprocal, where both mentor and mentee (aka protégé) derive emotional or tangible benefits; (4) is personal in nature, involving direct interaction; and (5) emphasizes the mentor’s greater experience, influence, and achievement within a particular organization.

The literature on mentoring within the health care field has run the gamut, from describing the value of mentoring in leadership, documenting a long-distance mentorship program, mentoring new faculty, using preceptors as mentors, and determining participation in mentoring relationships, to surveying the extent of administrative support for mentoring. Although much has been written on mentoring in health care, the research has not addressed the effectiveness of the mentoring relationship in the academic setting or the tools to measure that effectiveness.

Formal and informal mentoring programs have been popping up in colleges and universities nationwide, especially in medical schools such as ours. There are even a few books that describe guidelines for developing such programs. Unfortunately, criteria for evaluating the effectiveness of these programs are either not reported or not quantifiable. Within the context of this need, the Ad Hoc Faculty Mentoring Committee at Johns Hopkins University School of Nursing undertook to define the construct of “mentorship” and to develop new generic instruments to measure the effectiveness of a faculty mentoring relationship.

Interested faculty members in our school established the Ad Hoc Faculty Mentoring Committee to investigate faculty’s mentoring activities as they related to the criteria for promotion to associate professor and professor ranks. Although no formal mentoring program existed at our school, evidence of successful mentoring was required for promotion. That evidence was being submitted in the form of letters, written by faculty and student mentees, which described the mentoring relationship with the faculty candidate. There were no specific guidelines or criteria to define the mentoring relationship.
or to evaluate the mentor’s effectiveness. The evidence presented was nonstandardized and anecdotal.

The committee’s four tasks were to (1) define “mentorship,” (2) specify concrete characteristics and responsibilities of mentors that are measurable, (3) develop an instrument that provides a comprehensive profile description of the mentoring relationship, and (4) build a scale that measures the effectiveness of that relationship in terms of specific outcomes that a mentee could evaluate.

“Mentorship”: The Construct

The following products related to the definition and characteristics of mentorship are the results of the committee’s work.

Definition

Building on the previous definitions, but allowing for flexibility in the nature of the mentoring relationship, the committee proposed the following definition:

A mentoring relationship is one that may vary along a continuum from informal/short-term to formal/long-term in which faculty with useful experience, knowledge, skills, and/or wisdom offers advice, information, guidance, support, or opportunity to another faculty member or student for that individual’s professional development. (Note: This is a voluntary relationship initiated by the mentee.)

Characteristics

The desirable characteristics of a faculty mentor include, but are not limited to, expertise, professional integrity, honesty, accessibility, approachability, motivation, respect by peers in field, and supportive-ness and encouragement.

Responsibilities

In order to put some teeth into the role of mentor, faculty must commit to certain responsibilities for which he or she will be held accountable by the mentees. Those concrete responsibilities are:

- Provides timely, clear, and comprehensive feedback to mentee’s questions
- Respects mentee’s uniqueness and his or her contributions
- Appropriately acknowledges contributions of mentee
- Shares success and benefits of the products and activities with mentee

New Mentorship Effectiveness Instruments

Instruments related to mentoring programs in the 1980s relied on global questions about whether or not someone had a mentor\cite{38,40} or on a wide array of mentoring characteristics or functions.\cite{6,41–46} These tools mirrored the inconsistency in definitions and lack of consensus on a generic set of functions mentioned previously. In the 1990s, the health care research reporting instruments to measure the effectiveness of mentoring programs consisted of three studies,\cite{2,33,47} including only one in medicine.

Morzinski et al.\cite{47} described the evaluation of a formal mentoring program for junior faculty in academic family medicine. This program was based on skills deemed important for socialization in medicine and professional development. The mentees evaluated the impact of a mentoring program based on their achievement of three dimensions: development of career management skills, improved understanding of values and norms of the environment, and the ability to develop professional relationships. The study showed that junior faculty improved their professional and academic skills after participating in the mentoring program. The benefits were greater when the participants engaged in joint academic projects.

Critique of mentorship instruments

The current evaluation tools used in mentoring programs have several limitations in the context of the mentoring relationship. These tools are designed to evaluate only specific mentoring programs. They measure the importance of mentoring functions, and/or they measure the frequency of mentoring behaviors. And, these evaluation tools may or may not apply to faculty mentoring. Their questionnaire formats consist of short-answer constructed response and bipolar anchor scales with different anchors for each item. For example, the extreme anchors may be “Very Unsatisfied” and “Very Satisfied,” “Very Unimportant” and “Very Important,” “No Support” and “High Support,” and “No Impact” and “High Impact.” These bipolar anchors measure different characteristics about specific items. The scores on these items cannot be summed using a summated ratings procedure to produce subscale or scale scores. There are, however, other bipolar importance and activity scales with the same anchors. When anchors are presented only at the extreme ends of a scale continuum, the meaning of the scale’s values and the interpretation of responses at points in between these extremes are unclear or ambiguous. Only the respondents will know the true meaning of these points.

Despite all of the aspects of mentoring that these instruments measure, none measured the critical dimension, “To what extent were any of the relationships effective?” A rating scale that evaluates the degree of effectiveness was needed. The challenge was to develop such a scale, plus address the structural limitations of previous scales.

Given the variation and complexity of faculty mentoring relationships, measuring effectiveness seems inextricably linked to the nature of each unique relationship. Consequently the committee developed two instruments: a questionnaire that described the characteristics of the mentoring relationship (albeit a profile of the relationship) as seen from the perspective of the mentee, and a formal rating scale that measured the effectiveness of the mentor against the aforementioned characteristics and responsibilities.

The Mentorship Profile Questionnaire

The Mentorship Profile Questionnaire (see Appendix A) was developed to describe the exact nature of the mentoring relationship and to specify the outcome measures produced from the relationship. The Description Section requests the mentee define the role of his or her mentor (teacher, counselor, advisor, sponsor, advocate, resource), the frequency and mode of communication, length of the relationship, and its strengths and weaknesses. The Outcomes Section asks the mentee to identify, describe, and provide supporting documents for the products of the relationship, such as publications, presentations or posters, new teaching methods, clinical expertise, conducting research, service...
activities, job change or promotion, and grant writing.

The Mentorship Effectiveness Scale

Item generation. The committee constructed a formal rating scale to provide an efficient, comprehensive, and standardized tool for rating the mentorship experience and, especially, the effectiveness of the mentor (see Appendix B). Deriving the content from the pool of positive or desirable characteristics and responsibilities of mentors listed previously, 12 statements were generated to reflect a comprehensive assessment of the mentorship’s effectiveness. The statements were written to meet established scale-item criteria, gleaned from a variety of classic sources on scale construction. The items were then reviewed by a five-member faculty committee for their psychometric form as well as for their mentor-characteristic substance to provide evidence of content-related validity. Item revisions required several iterations until unanimity by the committee was attained.

Scale structure. A Likert-type summated rating scale format was used to elicit each mentee’s responses to the items. A highly discriminating six-point agree–disagree continuum was developed: 0 = Strongly Disagree, 1 = Disagree, 2 = Slightly Disagree, 3 = Slightly Agree, 4 = Agree, 5 = Strongly Agree. These anchors seemed most appropriate to evaluate responses to a wide range of mentors’ characteristics. No uncertain or neutral position was presented in order to force an agree–disagree rating. Nunnally and Bernstein have indicated there is a slight advantage to using an even-numbered scale with no middle “undecided” position because a neutral position response gives no rating information. A “Not Applicable” option was also listed in case a characteristic was not appropriate for a particular mentor–mentee relationship.

Response bias. Two types of response bias were of concern: acquiescence (or yea-saying) and the halo effect. Although these biases are not common with Likert-type scales, a mentee’s close working relationship with his or her mentor may affect the rater’s objectivity. The tendency to give positive responses to the “positive” characteristics, irrespective of the item content, or to rate the specific characteristics highly because of an overall positive impression of the mentor, can inflate the ratings and, consequently, favorably skew the responses. Given the nature of the mentor–mentee relationship, no psychometric antidote for this potential subjectivity and ratings bias appears possible. These effects should be considered in the interpretation of the final ratings.

Scale administration and scoring. Mentors nominate mentees to complete this scale. Each mentee rates the extent to which the mentor exhibited each of the 12 characteristics or met the behavioral descriptions. Degree of agreement represents a qualitative rating, albeit an ordinal score value, from which the mentor’s effectiveness could be inferred. The ratings may be presented item-by-item based on the 0–5-point quantitative scale or summed across all 12 items for a total rating, ranging from 0–60.

If several mentees rate the same mentor and the relationships are comparable, a median rating for the sample of mentees can be computed by item and for the total scale. This comparability, however, is quite rare. Most often, each mentor–mentee relationship is unique on one or more characteristics. This precludes aggregating ratings across mentees for a single mentor. If such ratings were combined, the results could be misleading and misrepresent the effectiveness of the mentoring.

Psychometric issues

Although there is considerable variation in the types of formal and informal mentoring programs in medical schools, minimal attention has been devoted to the development of instruments to evaluate mentors and the mentoring relationship. The Ad Hoc Faculty Mentoring Committee spent more than a year reviewing the literature and constructing the two tools described in this section. Despite the effort expended, there are built-in intractable psychometric issues that limit the collection of validity and reliability evidence. Such evidence is required by the Standards for Educational and Psychological Testing.

Content-related validity. The most important validity evidence is content related. The items on the Mentorship Profile Questionnaire and the Mentorship Effectiveness Scale must be congruent with the definition of mentoring and the domain of mentor characteristics and responsibilities in whatever mentoring activities occur. If there is a formal program, then the items must match the salient characteristics of the program as well. A panel of faculty members knowledgeable about mentorship should formally review the scale items against the mentoring characteristics and the panel should attain consensus or, preferably, unanimity.

Response bias. As noted previously, acquiescence and halo biases can inflate the ratings by the mentees. Although the direction of the bias can be anticipated, the degree cannot be measured. Either or both sources of response bias can lower the validity of the ratings, and the inferences drawn from them, about mentor effectiveness.

Other validity and reliability evidence. The most common indices of item analysis, validity, and reliability computed from sample data cannot be estimated for most scales of mentors’ effectiveness. Although a common set of criteria and scale items are administered using standardized procedures, typically each mentor–mentee relationship is unique. For example, the details of the relationships on the Mentorship Profile Questionnaire preclude the aggregation of ratings across mentees for the same mentor (see Appendix A). The ratings by each mentee are usually based on different role profiles. Hence, the ratings are not comparable and do not have the same meaning. Since a statistical sample of mentor ratings cannot be obtained, validity coefficients and standard indices of internal consistency reliability, such as coefficient alpha, as well as other group-based psychometric statistics, cannot be computed.

Conclusions

The research and experience on faculty mentoring relationships in academia, and medical schools in particular, over the past 25 years have produced lists of definitions, functions, and programs, but miniscule evidence of effectiveness. The concept of mentoring remains unclear and imprecise and instruments designed to evaluate mentoring programs rarely do. The effectiveness of formal and informal medical faculty mentoring programs intended to promote the professional growth of junior faculty and the academic success of students is based more on assumption than on demonstrated empirical evidence.
In view of this shaky foundation, the Ad Hoc Faculty Mentoring Committee at the Johns Hopkins University School of Nursing has contributed a generic definition, set of characteristics and responsibilities, and a mentorship profile and rating scale to measure faculty mentorship effectiveness. Although these products were developed in the absence of a formal mentoring program, the content, items, and instrument structure can be applied, or easily modified, to fit most informal as well as formal programs already in operation.

There is a critical need for research on mentoring that must address the definitional and conceptual issues plaguing this domain for years. Neither the empirical nor theoretical published research has kept pace with the development of mentoring programs. The scarcity of rating scales that directly measure characteristics of the mentoring relationship, essential to evaluate any program’s effectiveness, requires immediate attention.

Although the psychometric issues we have identified tend to limit quantification of results to person-, relationship-, and program-specific contexts, a deliberate effort should be devoted to tackling these scaling problems. Hopefully, our contribution will furnish a springboard, at least, for additional effort should be devoted to tackling these scaling problems.

References
Appendix A

Mentorship Profile Questionnaire Developed by the Ad Hoc Faculty Mentoring Committee, Johns Hopkins University School of Nursing

Your name: ____________ Mentor’s name: ____________

Part I: Description of Relationship

1. What was the role of your mentor? (e.g., teacher, counselor, advisor, sponsor, advocate, resource)
2. How often did you communicate? (e.g., e-mail, in person, telephone)
3. How long have you had this relationship?
4. How would you characterize the strengths and weaknesses of your relationship?

Part II: Outcome Measures

Directions: Please check all of the following that resulted from your interaction with your mentor and specify or describe below. Supporting documents may be attached, as appropriate.

1. ☐ Publication:
2. ☐ Presentation or poster:
3. ☐ New teaching method or strategy:
4. ☐ Clinical expertise:
5. ☐ Conducting research:
6. ☐ Service activities (e.g., community service, political activity, professional organization):
7. ☐ Development of a program (e.g., educational/clinical course or new program of study):
8. ☐ Job change/promotion:
9. ☐ Grant writing/submission:
10. ☐ Other:

Copyright © 2002 The Johns Hopkins University School of Nursing
Appendix B

Mentorship Effectiveness Scale Developed by the Ad Hoc Faculty Mentoring Committee, Johns Hopkins University School of Nursing

Your name: ____________________

Directions: The purpose of this scale is to evaluate the mentoring characteristics of ____________, who has identified you as an individual with whom he/she has had a professional, mentor/mentee relationship. Indicate the extent to which you agree or disagree with each statement listed below. Circle the number that corresponds to your response. Your responses will be kept confidential.

0 = Strongly Disagree (SD)
1 = Disagree (D)
2 = Slightly Disagree (SID)
3 = Slightly Agree (SlA)
4 = Agree (A)
5 = Strongly Agree (SA)
6 = Not Applicable (NA)

SAMPLE: My mentor was hilarious. 0 1 2 3 4 5

1. My mentor was accessible. 0 1 2 3 4 5 6
2. My mentor demonstrated professional integrity. 0 1 2 3 4 5 6
3. My mentor demonstrated content expertise in my area of need. 0 1 2 3 4 5 6
4. My mentor was approachable. 0 1 2 3 4 5 6
5. My mentor was supportive and encouraging. 0 1 2 3 4 5 6
6. My mentor provided constructive and useful critiques of my work. 0 1 2 3 4 5 6
7. My mentor motivated me to improve my work product. 0 1 2 3 4 5 6
8. My mentor was helpful in providing direction and guidance on professional issues (e.g., networking). 0 1 2 3 4 5 6
9. My mentor answered my questions satisfactorily (e.g., timely response, clear, comprehensive). 0 1 2 3 4 5 6
10. My mentor acknowledged my contributions appropriately (e.g., committee contributions, awards). 0 1 2 3 4 5 6
11. My mentor suggested appropriate resources (e.g., experts, electronic contacts, source materials). 0 1 2 3 4 5 6
12. My mentor challenged me to extend my abilities (e.g., risk taking, try a new professional activity, draft a section of an article). 0 1 2 3 4 5 6

Please make additional comments on the back of this sheet.

Copyright © 2002 The Johns Hopkins University School of Nursing

Academic Medicine, Vol. 80, No. 1 / January 2005